Deep Learning Applications in Chest Radiography and Computed Tomography

Citation for published version:
Lee, SM, Seo, JB, Yun, J, Cho, Y-H, Vogel-Claussen, J, Schiebler, ML, Gefter, WB, van Beek, E, Goo, JM, Lee, KS, Hatabu, H, Gee, J & Kim, N 2019, 'Deep Learning Applications in Chest Radiography and Computed Tomography: Current state of the Art', *Journal of Thoracic Imaging*, vol. 34, no. 2. https://doi.org/10.1097/RTI.0000000000000387

Digital Object Identifier (DOI):
10.1097/RTI.0000000000000387

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Thoracic Imaging

Journal of Thoracic Imaging

Deep Learning Applications in Chest Radiography and CT: Current State of the Art

--Manuscript Draft--

Manuscript Number:	JTI-18-112R1
Full Title:	Deep Learning Applications in Chest Radiography and CT: Current State of the Art
Article Type:	Review Article
Section/Category:	Pulmonary/thoracic
Keywords:	Chest imaging; Machine Learning; Deep Learning; Radiography; computed tomography; magnetic resonance imaging
Corresponding Author:	Joon Beom Seo, M.D. University of Ulsan College of Medicine Asan Medical Center Seoul, KOREA, REPUBLIC OF
Corresponding Author's Institution:	University of Ulsan College of Medicine Asan Medical Center
Order of Authors:	Sang Min Lee
	Joon Beom Seo, M.D.
	Jihye Yun
	Young-Hoon Cho
	Jens Vogel-Claussen
	Mark Schiebler
	Warren Gefter
	Edwin van Beek
	Jin Mo Goo
	Kyung Soo Lee
	Hiroto Hatabu
	James Gee
	Namkug Kim
Manuscript Region of Origin:	KOREA, REPUBLIC OF
Abstract:	Summary Deep learning is a genre of machine learning that allows computational models to learn representations of data with multiple levels of abstraction using numerous processing layers. A distinctive feature of deep learning, compared to conventional machine learning methods, is that it can generate appropriate models for tasks directly from the raw data, removing the need for human-led feature extraction. Medical images are particularly suited for deep learning applications. Deep learning techniques have already demonstrated high performance in detection of diabetic retinopathy on fundoscopic images and metastatic breast cancer cells on pathologic images. In radiology, deep learning has the opportunity to provide improved accuracy of image interpretation and diagnosis. Many groups are exploring the possibility of using deep learning based applications to solve unmet clinical needs. In chest imaging, there has been a large effort to develop and apply computer-aided detection (CAD) systems for the detection of lung nodules on chest radiographs and chest computed tomography. The essential limitation to CAD is an inability to learn from new information. To overcome these deficiencies, many groups have turned to deep learning approaches with promising results. In addition to nodule detection, interstitial lung disease recognition, lesion segmentation, diagnosis and patient outcomes have been addressed by deep learning approaches.

	The purpose of this review article is to cover the current state of the art for deep learning approaches and its limitations, and some of the potential impact on the field of radiology, with specific reference to chest imaging.

U. Joseph Schoepf, M.D.
Editor-in-Chief
Journal of Thoracic Imaging

June 1, 2018

Dear Dr. Schoepf:

We submit a manuscript entitled “Deep Learning Applications in Chest Imaging: Current State of the Art” for consideration of publication in Journal of Thoracic Imaging. This manuscript was invited to the IWPFI group for review article by Journal of Thoracic Imaging.

This manuscript addressed the current state of the art for deep learning approaches and its limitations, and some of the potential impact on the field of radiology, especially in chest imaging. We presented basic concepts and applications of deep learning such as classification, segmentation, and detection. In chest imaging, chest radiography and chest CT is the mainstay of discussion and computer-aided detection system using deep learning was focused on. Furthermore, image normalization, disease pattern classification, and patient outcome prediction based on deep learning was also dealt with.

Due to extensive search of literature, the number of references reached 94.

This work has not been published previously or submitted elsewhere for review. All authors have read and approved this manuscript and none of the authors have conflicts of interest or financial disclosures.

We thank you for your consideration of our manuscript and look forward to your decision concerning its suitability for publication.

Sincerely,

Joon Beom Seo, MD

Department of Radiology,
University of Ulsan College of Medicine, Asan Medical Center,
88 Olympic-ro 43 Gil, Songpa-gu,
Seoul 138-736, Korea.
E-mail: seojb@amc.seoul.kr

Our detailed responses to the recommendations and comments of the reviewers are as follows.

Reviewer #1:

I would like to congratulate the authors on this excellent review. This is a well-written, coprehensive, clinically relevant overview of the current status of deep learning in chest imaging. The manuscript combinds an excellent introduction into the technical aspects of deep learning with expert knowledge of clinical chest radiology - a rare and precious combination. The literature review is up to date. Figures are informative and of high quality. Limitations and challenges are discussed adequately.

I would suggest the authors comment in some more detail on two issues:

1-1. In clinical practice, the (differential) diagnosis is greatly dependent on factors other than the imaging features of the examination at hand: reasons for the examination, comparison with previous studies, clinical information such as medical history, immune status, laboratory parameters etc. This adds a much greater complexity to the task that deep learning algorithms need to master. Please comment on how such challenges are being addressed in deep learning.
→ Thank you for your comment. We also believe that the diagnosis should be made in the clinical context using various information and deep learning based clinical decision support system should integrate information in the different domains. Although many groups have focused on medical imaging at present, efforts to add clinical or pathological information using deep learning network and determine the output has been made. For example, Suk et al. showed a deep learning-based method for Alzheimer's Disease and Mild Cognitive Impairment diagnosis using multimodality information including CSF findings and Minimum Mental State Examination. We expect to see more advanced and flexible deep learning network dealing with various medial data in the near future.

Reference: Suk HI, Shen D. Deep learning-based feature representation for AD/MCI classification. Med Image Comput Comput Assist Interv. 2013;16:583-590.

1-2. Reference standard remains a major issue, especially in areas where the performance of human readers is poor (such as diagnosing pneumonia on a chest x-ray, distinguishing different insterstitial lung diseases on chest CT). Whenever claims are made that deep learning algorithm have shown superior performance to human readers (also in various places in this manuscript), it should be made more transparent what exactly was used as the reference standard to train the algorithms - and the validity of the reference standard should be critically analyzed.

→ We understand and agree with your concern on ground truth issue. As you mentioned, some abnormalities cannot be definitely diagnosed in imaging modalities. Therefore, subspecialty radiology societies can play important roles in defining appropriate tasks for deep learning algorithms as well as assisting in making publicly available strongly labeled training data sets and validation datasets.

In term of performance of deep learning algorithm, we also agree with your opinion that careful and thorough validation is required and the validity of the reference standard should also be critically analyzed.

Reviewer #2: This review presents an overview of deep learning (DL) algorithms applied to chest radiography and CT images.

2-1. As the main focus of this paper is on CT, I would suggest to change the title of the manuscript in: "Deep learning applications in chest radiography and CT: Current state of the art". Otherwise this title is in my opinion misleading the reader to expect a larger portion of e.g. MRI DL as well.

→ We followed your suggestion and made the appropriate change in the title page.

2-2. In this regard, I would suggest to delete section III. CMR completely, because this does not fit in the review and more importantly is lacking a lot of references on MRI DL.
I would also like to invite the authors to shortly comment about other radiological imaging technologies (MRI, PET, US) and their deep learning methods. In comparison to CT the other imaging technologies are way more challenging for DL applications because image content is not as consistent (image intensity, …) as in CT. Please also discuss this shortly.

→ We followed your suggestion and made the appropriate changes in pages 16, 17, 18, 31, and 32 of the annotated manuscript.

Specific comments:

2-3. Introduction: I do not fully agree with the statement "no a priori bias to the extraction", because there is a large bias depending on which architecture, loss function, training set etc. was chosen. The performance is at the moment a priori dependent on this defined setting. The authors commented on this part in the discussion, therefore I would rephrase this part. Otherwise readers might tend to think of DL as a magical tool which can generalize and perform any task. This should not be the statement of this review!

→ We followed your suggestion and made the appropriate change in page 3 of the annotated manuscript.

it can extract fully automated features and generate appropriate models for tasks directly from the raw data on its own, removing the need for human-led feature extraction~~. There is no a priori bias given to the extraction process except for the desired outcomes, such as separation~~

~~of image data into two or more groups.~~

2-4. Introduction, 2nd paragraph: Typo "medical field, the application …."

→ We followed your suggestion and made the appropriate change in page 3 of the annotated manuscript.

2-5. DL and CNN, 2nd paragraph: A CNN is not necessarily composed of convolutional, pooling and fully-connected layers. There exist CNNs with only convolution and softmax activation output. Please rephrase to "The architecture of a CNN can be composed..."

→ We followed your suggestion and made the appropriate change in page 5 of the annotated manuscript.

2-6. DL and CNN, 2nd paragraph: I do not understand why a pooling layer should retain the shape and position of the detected semantic feature. The sole purpose of a pooling layer is to reduce the input to the next stage: coarse-to-fine feature extraction and reduction in trainable parameters. Please correct it.

There are also other techniques besides pooling (dilated convolutions, convolutions with stride, …) which are worth to be mentioned.

→ We followed your suggestion and made the appropriate change in page 5 of the annotated manuscript.

2-7. DL and CNN, last paragraph: "However, there is still concern that deep learning is overhyped and that we still need rigorous clinical validation of this technology."

This statement reads at the moment as if rigorous clinical validation is not necessarily needed. I don't think that this was the intended message?! Please rephrase it.

→ We followed your suggestion and made the appropriate change in page 4 of the annotated manuscript.

2-8. Classification: The authors should better clarify that at the moment there are several different "building blocks" available which were introduced by the respective architectures (VGG, ResNet, Vnet, ...): conv, pooling, BN, inception, atrous conv, dilated conv, residual, dense, …). Current DL architecture designs use combinations of these building blocks to perform the desired tasks - instead of relying on the bare previously proposed architecture.

→ We followed your suggestion and made the appropriate change in page 6 of the annotated manuscript.

2-9. Perspective, challenges, …: There have also been other work which provided a feedback to the reader: Kuestner et al. MRI 2018 (10.1016/j.mri.2018.07.003), Lorente et al. ISBI 2014 (10.1109/ISBI.2014.6868128). Please cite them as well.

→ Thank you for your suggestion, but, we are afraid that the two articles which you recommended are out of our scope.

Deep Learning Applications in Chest Radiography and CT: Current State of the Art

Sang Min Lee, MD[1]; Joon Beom Seo, MD[1]; Jihye Yun, PhD[2]; Young-Hoon Cho, MD[1]; Jens Vogel-Claussen, MD[3]; Mark L. Schiebler, MD[4]; Warren B. Gefter, MD[5]; Edwin J.R. van Beek, MD[6]; Jin Mo Goo, MD[7]; Kyung Soo Lee, MD[8]; Hiroto Hatabu, MD[9]; James Gee, PhD[10]; Namkug Kim, PhD[1,2]

[1]Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea

[2]Department of Convergence Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, South Korea

[3]Institute of Diagnostic and Interventional Radiology, German Center for Lung Research, Hannover Medical School, Hannover, Germany

[4] Department of Radiology, University of Wisconsin-Madison School of Medicine and Public Health, Madison, WI, USA

[5]Department of Radiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

[6]Department of Radiology, University of Edinburgh, and Edinburgh Imaging, Queen's Medical Research Institute, Edinburgh, Scotland, UK

[7]Department of Radiology, Seoul National University College of Medicine, and Institute of Radiation Medicine, Seoul National University Medical Research Center, Seoul, Korea

[8]Department of Radiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

[9]Department of Radiology, Brigham and Women's Hospital and Harvard Medical School, Boston, USA

[10]Penn Image Computing and Science Laboratory, Department of Radiology, University of Pennsylvania, Philadelphia, USA

Running head: Deep Learning Applications in Chest Imaging

Corresponding Author: Joon Beom Seo

Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, 88 Olympic-ro 43 Gil, Songpa-gu, Seoul 138-736, Korea Tel: 82-2- 3010-4400; Fax: 82-2- 476-4719; E-mail: seojb@amc.seoul.kr

Conflict of interest statements: All authors have no conflicts to disclose.
(Edwin J.R. van Beek discloses the following: Advisory boards: Imbio, Aidence; Owner/founder: Quantitative Clinical Trials Imaging Services)

Deep Learning Applications in Chest ~~Imaging~~Radiography and CT: Current State of the Art

Commented [이상1]: R2-1

Type of Manuscript: Invited review

Summary

Deep learning is a genre of machine learning that allows computational models to learn representations of data with multiple levels of abstraction using numerous processing layers. A distinctive feature of deep learning, compared to conventional machine learning methods, is that it can generate appropriate models for tasks directly from the raw data, removing the need for human-led feature extraction.

Medical images are particularly suited for deep learning applications. Deep learning techniques have already demonstrated high performance in detection of diabetic retinopathy on fundoscopic images and metastatic breast cancer cells on pathologic images. In radiology, deep learning has the opportunity to provide improved accuracy of image interpretation and diagnosis. Many groups are exploring the possibility of using deep learning based applications to solve unmet clinical needs.

In chest imaging, there has been a large effort to develop and apply computer-aided detection (CAD) systems for the detection of lung nodules on chest radiographs and chest computed tomography. The essential limitation to CAD is an inability to learn from new information. To overcome these deficiencies, many groups have turned to deep learning approaches with promising results. In addition to nodule detection, interstitial lung disease recognition, lesion segmentation, diagnosis and patient outcomes have been addressed by deep learning approaches.

The purpose of this review article is to cover the current state of the art for deep learning approaches and its limitations, and some of the potential impact on the field of radiology, with specific reference to chest imaging.

Keywords

Chest imaging; Machine learning; Deep learning; Radiography; Computed tomography; Magnetic resonance imaging

Introduction

Deep learning is a genre of machine learning that allows computational models to learn representations of data with multiple levels of abstraction through the use of a number of unique processing layers [1]. The most distinctive feature of deep learning, compared to the conventional machine learning methods, is that it can extract fully automated features and generate appropriate models for tasks directly from the raw data on its own, removing the need for human-led feature extraction~~. There is no a priori bias given to the extraction process except for the desired outcomes, such as separation of image data into two or more groups~~.

Commented [이상2]: R2-3

In recent years, deep learning methods have shown breakthroughs in various fields including image recognition [2], speech recognition [3] as well as information technology. However, in the ~~medial~~ medical field, the application of deep learning is currently in its infancy. Medical images and their respective patient electronic medical records are well suited for analysis by deep learning. Some of the first successful demonstrations of deep learning techniques were reported in the detection of lymph node metastasis from hematoxylin and eosin stained pathologic micrographs, analysis of skin cancer from photographs of the lesion, and the diagnosis of diabetic retinopathy from fundoscopic images [4-7].

Commented [이상3]: R2-4

In radiology, deep learning will help to improve efficiency by automated image interpretation and generation of an appropriate differential diagnosis. Data mining of the patient's electronic medical record data (big data) combined with deep learning applied to the patient's medical images should help to improve patient outcomes. Cloud based applications also allow the deep learning algorithm to continuously learn on data sets that are not restricted to a single institution. Many groups are now exploring deep learning-based applications for solutions to unmet clinical needs. In chest imaging, significant effort has been directed at developing and applying computer-aided detection (CAD) [8] systems for the detection of nodules on chest radiographs and chest computed tomography (CT) [9, 10]. Although many CAD systems are being used in clinical practice, the implementation of CAD has not been widely accepted due to its poor performance (i.e. frequent false positive and false negative cases). Deep learning approaches have the potential to overcome the limitations of existing CAD systems, with several studies showing promising results [11, 12]. Moreover, disease pattern recognition, lesion segmentation, diagnosis and survival prediction have been successfully studied using deep learning in chest imaging [13]. However, there is still concern on this technology in terms of clinical application~~that deep learning is overhyped and that we still need rigorous clinical validation of this technology~~.

In this review article, we introduce the principle methods of deep learning, their potential applications and clinical promise in chest imaging.

Commented [이상4]: R2-7

Deep Learning and Convolutional Neural Networks

Machine learning is defined as a set of methods that can automatically detect patterns in data, and then utilize the uncovered patterns to classify, predict, or conduct various types of decision making under uncertain conditions [14]. Conventional machine learning techniques rely on extensive data engineering and considerable domain expertise to design a “feature extractor” algorithm that converts the raw data into suitable representations for computational analysis. A convolutional neural network (CNN) is a special type of deep learning and is quite similar to the overall learning process (e.g. neuronal pruning) of the mammalian visual cortex [15], and is responsible for the recent improvements in the field of computer vision (e.g. self-driving automobiles). With the availability of large datasets and increased computing power, CNNs have produced promising results for many tasks including image classification, correct image detection, correct image segmentation and understanding speech (e.g. natural language processing).

The architecture of ~~a typical CNN is~~a CNN can be composed of convolutional, pooling, and fully connected layers (Figure 1): (1) the convolutional layers detect distinctive local motifs by applying multiple filters and generating multiple feature maps, (2) the pooling layers effectively reduce the dimensions of feature maps (other techniques such as dilated convolutions, convolutions with stride can also be used)~~, and robustly retains the shape and position of detected semantic features within the image~~, (3) the fully connected layers integrate all feature responses and eventually project onto an output layer which serves to answer the task at hand. By using deep CNN architecture (repeating the convolutional and pooling layers several times) to mimic the natural neuromorphic multi-layer network, deep learning can automatically and adaptively learn a hierarchical representation of patterns and consequently identify the most significant features for a given task [2]. In order to deal with complex tasks, networks with many layers – so called Deep Networks – are required. However, adding additional layers increases the number of parameters in the model and can make it more difficult to train it for a specific task without overfitting the data.

Commented [이상5]: R2-5

Commented [이상6]: R2-6

1. Classification

One key task for radiologists is creating an appropriate differential diagnosis for each patient's medical images. This job can be computationally defined as a typical classification task using input from medical images and any available clinical information. There are many different CNN network architectures for classifying images. In order to improve the efficiency of the training procedure and reduce the number of parameters, the deeper networks have introduced more effective subroutines - "building blocks". These building blocks are small branching/spanning convolution blocks with pooling and batch normalization layers, which can be repeated to construct deeper architectures [16]. VGG19 [17] used small and fixed size kernels in each layer to win the ImageNet challenge of 2014. Another CNN named GoogLeNet (i.e. " Inception") [18] made use of the building block that is a multi-level feature extractor with a set of convolutions of different sizes. ResNet [19], which won the ImageNet challenge of 2015, introduced the subroutine of a "residual building block" which was designed to learn the residual (e.g. features that remain important) in order to make it easier to train deeper neural networks. This residual block was implemented by adding the input of the block to the output of the layers within the block (Figure 2). Since 2014, the performance on the ImageNet benchmark has saturated, but use of these architectures remains popular for medical image processing.

Moreover, these days, various combinations of these building blocks are used to construct deep learning architectures for the desired tasks, instead of relying on the bare previously proposed architecture.

Commented [이상7]: R2-8

2. Segmentation

The innovations of object classification have now shifted to semantic segmentation. This is a common task for both natural and medical image analysis whereby each voxel is classified in an image to determine the boundary conditions that define a specific object. The fully convolutional network (FCN) [20] represents a critical breakthrough for deep learning based semantic segmentation. In a FCN, the fully connected layers in the standard CNNs are replaced by convolutions with large receptive fields. This method achieves this segmentation by using coarse class score maps obtained by feed forwarding an input image. U-Net [21], which is the most well-known segmentation architecture in medical image analysis, combines an equal amount of upsampling and downsampling layers with skip connections between opposing convolution and deconvolution layers. Mask Region-based CNN (R-CNN) [22] detects objects in an image while simultaneously generating segmentation masks for each instance. This method has achieved the state-of-the-art performance on Microsoft Common Objects in Context [23].

3. Detection

The detection of objects of interest (i.e. lesions) is a key part of diagnosis and is one of the most labor-intensive tasks for radiologists. Several CNN network architectures have been shown to be able to detect a variety of objects quickly and accurately. R-CNN [24] combines region proposals (from a defined set of candidate detections) with CNNs and has then been improved to Fast R-CNN [25] and Faster R-CNN [26] with better performance. There are a few methods that computationally approach the problem of image detection by using multivariate regression; two of the most popular CNNs are You only look once (YOLO) [27] and Single shot multibox detector (SSD) [28], successfully predicting bounding boxes and classification probability.

4. Generative Model

Generative adversarial networks (GANs) have the advantage of automatically producing new images (e.g. synthetic image data) similar to samples from the training set by using two competing CNNs where one is generating artificial samples and the other is discriminating artificial from real samples [29] (Figure 3). These GANs could be trained end-to-end and learn representative features in a completely unsupervised manner. The representations learned by GANs are employed in various applications including medical image syntheses [30, 31], image normalization [32, 33], and super-resolution [34, 35].

Applications in Chest Imaging

I. Chest Radiography

Chest radiography is the most commonly performed diagnostic imaging procedure and over 35 million chest radiographs are performed each year in the United States alone and the average radiologist reads more than 100 chest radiograph exams per day [36]. Although these exams are clinically useful, efficient, and cost-effective, chest radiography consists of complex 3-dimensional anatomic information condensed in a 2-dimensional projection. Accurate interpretation of chest radiographs requires a great deal of experience and medical knowledge on the part of the radiologist. Increased radiologist work-loads combined with the intrinsic challenges of interpreting chest radiographs is associated with considerable inter- and intra-reader variability, missed lesions, and reporting delays in today's medical practice [8]. Deep learning technology has the potential to automatically detect abnormalities or assist radiologists in reading chest radiographs. Such technology would be very attractive for rural areas with few radiologists as well as for state-of-the-art medical centers to help support high volume workflows and improve efficiency of the radiology departments [12].

1. Lung Nodule Detection

Lung nodule detection from chest radiographs is another promising area for the application of deep learning technology. Lung cancer is the leading cause of cancer death worldwide and chest radiography has been the most widely adopted screening and imaging tool to detect lung cancer. However, unfortunately, due to the confounding effects of anatomic complexity on chest radiographs, lung cancer screening using plain chest film has yielded unsatisfying results, with reports of missed nodules being as high as 40% [37, 38].

CAD systems have been developed to help radiologists detect lung nodules. Recently this method has shown a sensitivity of 71% with 1.3 false-positive CAD marks per image[39]. Using bone-suppressed dual energy chest radiographs, another stand-alone

CAD system achieved a sensitivity of 74% with a 1.0 false-positive CAD mark per image [9]. In the setting of follow-up of patients with previous cancer (of any type), another CAD system showed promise by improving sensitivity from 63% to 92% for detection of lung nodules while only slightly decreasing specificity (from 98% to 96%) [40]. Although there has been improvement in CAD nodule detection on chest radiograph, these methods still need better accuracy before they are routinely accepted.

Recently CAD systems using deep learning techniques have shown improved accuracy for nodule detection on chest radiograph. A deep learning based technique identified by Wang et al. extracted deep learning features by transfer learning and combined them with traditional hand-crafted features. This CAD system achieved a higher sensitivity (69.3%) for nodule detection at a significantly lower false positive rate (1.19 false positive marks per exam) [41]. There is a recent report using CNN with visual attention networks generating respective accuracies of 0.76 for nodule detection and 0.65 for nodule localization on chest radiographs [42].

2. Diagnosis of Tuberculosis

Another specific field of research with great potential benefit to public health is utilizing deep learning technology for the diagnosis of pulmonary mycobacterium tuberculosis (TB) based on chest radiography. TB is an infectious disease caused by the acid fast (i.e. outer capsule appears red on hematoxylin and eosin staining) bacillus mycobacterium tuberculosis. In western countries, it is often thought as a "disease of the past", but TB is still a major health problem in the developing world, with millions of new cases encountered every year.

While the diagnosis of TB can be confirmed by bacteriology or the whole blood gamma interferon release assay (Quantiferon Gold, Qiagen), chest radiography is a highly sensitive imaging tool for triaging and screening for current and previous pulmonary TB infection. This organism is very difficult to culture in vitro and is often not able to be confirmed bacteriologically [43]. In locations where the prevalence of TB is high, there are a limited number of experienced chest radiologists available to use chest radiography as a method to confirm this disease. This shortage impairs screening efficacy and limits the opportunity to start medical therapy for a complete recovery [44]. Therefore, considerable effort developing CAD systems for the detection of pulmonary TB on chest radiographs has been extended. Traditional CAD systems without deep learning technology have shown acceptable TB detection performance with an area under the curve ranging from 0.71 to 0.84 [45].

Recently Lakhani et al. reported the performance of CAD using CNN for detection of pulmonary TB [12] and in that study, the CAD system reached an AUC of 0.99, which is greater than any previously reported CAD system. Although external validation is still needed to determine the true clinical benefit, CNN based CAD is a feasible and promising approach in this clinical scenario.

3. Multiple Abnormal Pattern detection

Although the detection of lung nodules and TB on chest radiographs has gained attention for many deep learning researchers, these findings can be relatively rare. Each chest radiograph may contain many abnormalities, for example: pneumonia, pleural effusions, pneumothorax, medical devices and cardiomegaly (Figure 4, 5). Therefore, facilitating deep learning technology to detect multiple abnormal patterns (MAP), rather than concentrating simply on nodules or TB would be more clinically practical.

Emergence of deep learning has drastically improved the performance of machine learning for object recognition, detection, and localization when compared to previous methodologies. Critical to the success of these methods are well-annotated ("strongly labeled") large datasets for effective system training. Recently two large datasets of chest radiographs, Open-I and ChestX-ray14, consisting of more than 110,000 chest radiographs from 30,805 patients, have been publicly released and have attracted considerable attention in the deep learning community. These publicly available data are external validation sets for any deep learning application using chest radiographs

In 2017, Wang et al. trained various known CNN models to detect 8 abnormal patterns (atelectasis, cardiomegaly, effusion, infiltration, mass, nodules, pneumonia, pneumothorax) on chest radiographs and achieved accuracy ranging from 0.56 to 0.78 [46]. Another study by Cicero et al. reported that a retrospective analysis of 35,038 chest radiographs from a single medical center using CNN (GoogLeNet), they were able to obtain MAP classification accuracy of 0.88 [11]. MAP detection in CXR with deep learning technology is still an area of active ongoing research and different methodologies are being tested and validated, and overall accuracy will likely improve.

II. Chest Computed Tomography

Unlike chest radiographs, chest CT provides cross-sectional images, allowing for direct 3-dimensional visualization of anatomic structures. Chest CT has a much higher sensitivity and lower inter-reader variability for detection of lung abnormalities and is frequently employed in the diagnosis and follow-up of most pulmonary diseases. Additionally, enhanced clinical availability, decreased cost, reduced radiation dose, and overall technical improvements of CT machines have resulted in a progressive increase in numbers of CT exams performed each year. Therefore, effective CAD system for chest CT interpretation would promote overall workflow for radiologists, by reducing the time required to read each CT exam and enhance reading accuracy.

1. Nodule Detection / Screening

Accurate nodule detection on chest CT has become a recent point of emphasis for efficient lung cancer screening. Despite advances in cancer treatment and screening programs, most lung cancer patients are still initially diagnosed at an advanced stage of the disease, which is associated with less than 20% 5-year survival [47].

Since the National Lung Screening Trial (NLST) announced a significant improvement (20%) in lung cancer mortality in high-risk populations when screened with low-dose chest CT (LDCT) [48], LDCT for cancer screening has been widely accepted [49]. Potentially this will lead to an increased volume of LDCT, which will require expert analysis from a radiologist for the detection and classification of nodules into either benign or malignant diagnoses.

A CAD system could aid radiologists in both detection and classification of lung nodules (Figure 6). Although traditional CAD systems have provided solid results, they often consist of complex pipelines of algorithms that depend heavily on manual human input such as pre-processing, segmentation, feature extraction, and model training, potentially hindering their performance [50]. Application of deep learning technology, on the other hand, can potentially remove innate challenges in traditional CAD systems by providing seamless feature identification and classification and removing the need for complex human-led feature extraction pipelines.

In 2011, the Lung Image Database Consortium (LIDC) database, containing 1018 cases of thoracic CT scans and image annotations by 4 thoracic radiologists, was released and has motivated deep learning researchers to develop CAD systems for chest CT nodule detection and classification [51]. CNNs are the most commonly utilized deep learning technology for lung nodule detection on CT images, and achieves good nodule detection sensitivity while maintaining an acceptable false positive rate. The first report of CAD system with deep learning technology for lung nodule detection on CT was Hua et al. in 2015, achieving sensitivity of 73% and specificity of 80%, which was superior to any other available conventional CAD systems [52]. In 2016, Setio et al. trained CNN to detect pulmonary nodules and achieved 85.4% sensitivity with only one false positive lesion per scan [53]. Studies that are more recent have shown the ability of CNNs to boost nodule detection sensitivity on CT to a higher level (95%) but were associated with a wide range (1.17 - 22.4) of false positive rates [54-56].

Classification of detected lung nodules is also a potential area that could benefit from the use of CAD systems. CT characteristics of a lung nodule, mainly nodule type and size, are closely associated with the likelihood of malignancy. These CT features are important determinants for planning treatment and follow up strategy. However, there is considerable observer variability in classification of pulmonary nodules among radiologists and this can lead to redundant follow-up examinations, unnecessary invasive procedures, or neglected malignancy [57]. In 2017, Ciompi et al. introduced a deep learning system that achieved good performance for nodule type classification based on lung-RADS system and was even within the inter-observer variability among four experienced human readers [58]. Furthermore, one study found that nodule classification accuracy of CAD system was improved by combining deep residual learning, curriculum learning, and transfer learning [59]. Other studies using different CNN models have achieved a classification accuracy as high as 87.1% [60, 61].

2. Interstitial Lung Disease

Interstitial lung disease (ILD) pattern classification is another area of research for deep learning technology. ILD is characterized by progressive fibrosis or inflammation of lung tissue and eventual deterioration of respiratory function [62]. Accurate diagnosis of ILD presents a challenge for the multidisciplinary medical panel at each institution that cares for these disorders because most ILDs have similar clinical manifestations, despite the fact that they are a histologically heterogeneous group of diseases with distinct prognoses.

High-resolution CT (HRCT) is currently the diagnostic imaging tool of choice for the diagnosis and evaluation of ILDs. However, ILDs have similar appearance on CT and CT readings are prone to high inter- and intra-observer variability [63]. Therefore, automatic identification and classification of different ILD patterns on chest CT may be helpful even for experienced chest radiologists, and application of deep learning technology could play an eminent role in developing such CAD systems. Segmentation of the lung with ILD could be enhanced by semantic segmentation with CNN [64]. In 2016, deep learning technology with CNN showed accuracy of 85% for classifying 6 different ILD patterns in dataset of 14,696 image patches [65]. In 2017, Kim et al. compared shallow and deep learning methods on classifying six ILD patterns on CT and found that deep learning methods showed significantly better accuracy and that accuracy was further increased with addition of more convolution layers [66] (Figure 7). More recently, a new CNN method achieved an ILD pattern classification accuracy of 87.9% using the holistic input of the entire CT data set [67]. Moreover, CAD methodology demonstrated a prognostic ability of lung function decline using quantifiable ILD on CT studies [68].

3. Chronic Obstructive Pulmonary Disease

A more basic field of application for deep learning technology is the segmentation and reconstruction of organs-of-interest from chest CT scans. Organ segmentation usually is the first step of many CAD systems, even those using deep learning methods, and accuracy of segmentation process is critical because any errors in this process would affect all the subsequent analysis. Various methods for organ segmentation have been developed and tested, showing promising results, but deep-learning based models could potentially improve methodological robustness and generalizability across imaging platforms, thus providing outcomes that are more reliable.

In 2017, Harrison et al. developed a deep model called progressive holistically-nested networks (P-HNNs) and reported that their P-HNNs model showed significant improvements of lung segmentation performance compared to previous segmentation approaches [69]. As for lobar segmentation, traditional methods are semi-automatic at best and largely relied on airway or vessel anatomy to delineate the lobar borders, with only few exceptions [70]. To address these problems, a deep learning method for lobe segmentation was introduced in 2017 and this method achieved high accuracy without reliance on prior airway or vessel segmentations, even when tested in lungs that had underlying disease [71, 72] (Figure 8).

Aside from lung tissue segmentation, robust and reliable airway segmentation is also essential for quantitative evaluation of various diseases involving the airways, such as chronic obstructive pulmonary disease (Figure 9). A large number of prior methods have common limitations that they are substantially influenced by morphologic changes in airway trees and measurement errors, such as airway leak that are most prevalent at smaller (or more peripheral) airways [73]. In fact, 15 different traditional algorithms were evaluated at an airway segmentation challenge in 2009 (EXACT 09), and precise delineation of a small bronchus without airway measurement leaks remained a common unsolved problem from this challenge [74]. In 2017, a deep learning method was developed and tested using a dataset from EXACT 09, and found that CNN significantly decreased airway leaks during segmentation process, resulting in higher sensitivity and specificity compared to all the other algorithms that participated in the EXACT 09 challenge [75]. In another study, even with incompletely annotated data, 3D deep fully convolutional networks demonstrated considerable improvements in airway segmentation while maintaining acceptable quantity of airway leaks [76].

4. Image Normalization

The reconstruction kernel is one of the most important technical parameters that determine the trade-off between spatial resolution and image noise in CT [77]. Since the selection of kernel affects quantitative analysis [78], CT images with different reconstruction kernels are necessary for various diagnostic or quantitative purposes. To overcome the limitation that it is difficult to save the raw data before reconstruction with various kernels, post-processing techniques have been developed to permit interconversion among CT images obtained with different kernels. Kim et al. [79] recently demonstrated that CNNs could be taught differences between high- and low-resolution images (residual images) and then they could be used to accurately and rapidly convert low-resolution images to high-resolution images. This approach is also applicable for interconverting CT images obtained using different kernels (Figure 10).

5. Radiomics and Deep Survival

Radiomics and prediction of patient outcomes (a.k.a." deep survival") are also active areas of research for the application of deep learning technology. Radiomics, which has gained substantial interest from researchers around the globe, involves the high-throughput extraction of quantitative features from medical images to develop reliable models to predict genomic information, clinical outcomes, and survival [80]. Extraction of radiomics features is a critical process in radiomics research and the majority of previous studies use handcrafted features, which are limited by current medical knowledge and human observation. On the other hand, CNN and transfer learning can be incorporated into radiomics models to extract more diverse features (deep features), which are free from prerequisite medical knowledge and biases. In this context, Lao et al. extracted 98,304 deep features (this would qualify as an example of over-fitting of the data) from images of glioblastoma multiforme and found 6 deep features that could predict overall survival with a concordance index of 0.71 [81].

In chest imaging, Paul et al. combined deep features of lung nodules detected on chest CT with traditional radiomics features to predict the probability of a malignant nodule and reported an overall accuracy of 76.8% and an AUC of 0.87 [82]. Another group used CNN to predict patient outcomes in a large cohorts of smokers and COPD patients and the CNN model predicted mortality with fair discrimination [13]. Deep radiomics and deep survival are promising new fields for study.

~~III. Cardiac Function Assessment~~

~~Cardiovascular magnetic resonance (CMR)~~

~~Cardiac mass and function are key parameters for diagnosis, monitoring and prognosis for numerous cardiac pathologies in clinical practice [83]. Currently the clinical workflow for reading CMR studies is prolonged by the time consuming semi-automated extraction of biventricular mass and function assessments, which require user interaction. Therefore, the human expert is still a factor for variability of cardiac function measurements especially between observers with different segmentation styles in multi-center multi-vendor settings [84]. Novel deep learning CNNs are expected to provide the potential to overcome human variability and reduce the time required for analysis.~~

~~The fully automated measurement of cardiac metrics from CMR exams is a particularly attractive problem, but previous attempts have been confounded by endocardial or epicardial border-tracking difficulties and have proved to be inaccurate [85, 86]. Novel machine learning approaches [87-89] are among the most successful methods for automated image segmentation of 2D short axis cardiac cine MRI stacks. Recently, Winther et al. developed v-net CNN, which achieved equal or better performance compared to the human expert readers in a multicenter multivendor data set [87]. Especially in the~~

~~anatomically complex right ventricle, the performance of v-net was superior (Figure 11). Also they could show that systematic differences between observers could be adjusted to account for different segmentation styles [90]. Using the fact that the prediction of a segmentation on one slice is dependent upon the already existing segmentation of an adjacent slice, Zheng et al. also achieved comparable or even better than the state-of-the-art results with data from the UK Biobank and three other cohorts [91]. Their method combines the strengths of 2D CNN methods but still addresses key 3D issues. Currently, high quality data from human experts (Strongly labeled and/or ground truth) is still needed to adequately train CNNs for the automated generation of CMR cardiovascular metrics. While guidelines already exist, further harmonization of segmentation methods between human experts and different sites are needed [92].~~

~~As CMR analysis software companies become more heavily invested in artificial intelligence, machine-learning approaches will begin to enter routine clinical application. These post processing tools will minimize the need for direct expert reader manipulation and thereby drastically reduce physician time to interpret these exams. The expert reader will simply be able to conduct a quality assessment of the fully automated segmentation results and interpret the functional cardiac MRI parameters in their appropriate clinical context. In the future, deep learning algorithms will automatically extract regional wall motion/strain, infarct size as well as T1 and T2 relaxation times, which are often not assessed due to time constraints.~~

Commented [이상8]: R2-2

Perspective, Challenges, and Limitations

In this manuscript, we reviewed basic concepts of deep learning and its various applications in chest radiography and CT~~imaging~~. In comparison to CT, MRI is more challenging for deep learning application because there is no pulse sequence dependent standardized intensity scale like the Hounsfield units in CT [93]. The application of this new technology to radiology has barely started, but it has shown remarkable results when compared with previous studies. We believe that these improvements in performance will soon offer new possibilities for the clinical practice of radiology.

Commented [이상9]: R2-1

Commented [이상10]: R2-2

The first deep learning based CAD application may be used to find critical findings on chest radiograph and triage the worklist before a radiologist's read. In brain CT, Prevedello et al. [94] already demonstrated that deep learning based algorithm could automatically identify critical findings and notify the interpreting radiologist. Furthermore, if the performance of CAD can be clinically acceptable in terms of prioritization of chest radiograph, it implies that deep learning based CAD have potential to differentiate normal chest radiographs from grossly abnormal exams. Thus, deep learning based CAD should improve the workflow and efficiency of radiology departments.

Second, CAD can help diagnosis of disease such as ILD and generate a preliminary quantitative report based on CAD results. This CAD report is repeatable with the same results and has no "intra-reader" variability. CAD combined with big data technology may retrieve similar images or diagnosis when radiologists require during interpretation of CT. It can also help to reduce the reading time.

Third, automation of lesion detection, segmentation, quantification by deep learning techniques facilitates reporting of the quantitative analysis of medical images more easily. Deep learning based segmentation tool improves accuracy and decreases image interpretation time. Furthermore, these data will likely provide improve the prediction of patient outcomes and risk stratification.

However, there are still many challenges to overcome. Currently, training deep learning algorithms requires large, strongly labeled and anonymized image datasets. These data sets are very challenging to acquire. While some abnormalities such as pneumothorax and malpositioned lines/tubes can be based upon imaging findings alone, most diseases require clinical documentation and/or pathological confirmation. Ambiguous or overlapping radiographic terms such as "consolidation" and "infiltrate" should not be used as surrogates for pneumonia to label training cases. This has been recognized as a limitation of some publically available datasets. National organizations (e.g. ACR, RSNA) and subspecialty radiology societies can play important roles in defining appropriate tasks for deep learning algorithms, as well as assisting in making publicly available strongly labeled training data sets and validation data sets.

Futhermore, the challenges regarding the ethical and legal aspects of data sharing and patient privacy are also paramount. There are severe monetary penalties (i.e. fines) in the United States of America [95] for any medical facility that allows compromise of personal health information/images. In the USA, the Health Insurance Portability and Accountability Act (HIPPA) governs any use of a patient's health information; as such, it is of critical importance that these imaging and medical data that are used for training, testing and validation of deep learning methods are fully anonymized and comply with this law. New data protection laws have also been introduced throughout Europe. As deep learning requires an enormous amount of high-quality data, the laws governing the safe handling of medical images and medical record data need to be followed. New technology, such as Blockchain, may be helpful in guarantying secure data sharing.

Lastly, we should demand a thorough and systematic clinical validation of any deep learning based applications as a prerequisite to commercial application. A well-known problem with these methods is overfitting and lack of utility when asked to analyze other data sets (e.g. poor interoperability). Most machine learning publications have shown their results in carefully preselected and enriched test sets (e.g. spiked to favor that algorithm with a higher prevalence of the condition than is found clinically). Thus, beyond just determining the feasibility of using any deep learning application in a test set chosen by the author, each deep learning application should be tested by a publicly available external validation set. We believe that this should be a requirement for any commercially approved deep learning method.

Conclusion

The application of deep learning methodology to help solving many tasks associated with medical imaging is at its infancy. While there are problems with every disruptive technological innovation, we believe that deep learning will soon be an indispensable tool for radiology. This is analogous to how the picture archiving communication systems and radiology information systems have transformed medical imaging and improved radiology while at the same time decreasing the cost of medical care. Reasonable expectations for this disruptive technology are needed, along with careful attention to any ethical, legal, and regulatory issues that may arise. This technology will enable radiologists to become more productive and improve patient care. The full potential of this technology will require radiologists to have an active role in governing its successful introduction to the clinic.

References

1. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436-444.
2. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012. 1097-1105.
3. Hinton G, Deng L, Yu D, et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. IEEE Signal Processing Magazine. 2012;29:82-97.
4. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA. 2016;316:2402-2410.
5. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542:115-118.
6. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA. 2017;318:2199-2210.
7. Ting DSW, Cheung CY, Lim G, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. JAMA. 2017;318:2211-2223.
8. McAdams HP, Samei E, Dobbins J, 3rd, et al. Recent advances in chest radiography. Radiology. 2006;241:663-683.
9. Schalekamp S, van Ginneken B, Koedam E, et al. Computer-aided detection improves detection of pulmonary nodules in chest radiographs beyond the support by bone-suppressed images. Radiology. 2014;272:252-261.
10. Liang M, Tang W, Xu DM, et al. Low-Dose CT Screening for Lung Cancer: Computer-aided Detection of Missed Lung Cancers. Radiology. 2016;281:279-288.
11. Cicero M, Bilbily A, Colak E, et al. Training and Validating a Deep Convolutional Neural Network for Computer-Aided Detection and Classification of Abnormalities on Frontal Chest Radiographs. Invest Radiol. 2017;52:281-287.

Formatted: Line spacing: Double

Field Code Changed

Formatted: Font: Verdana

12. Lakhani P, Sundaram B**.** Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology. 2017;284:574-582.
13. Gonzalez G, Ash SY, Vegas-Sanchez-Ferrero G, et al. Disease Staging and Prognosis in Smokers Using Deep Learning in Chest Computed Tomography. Am J Respir Crit Care Med. 2018;197:193-203.
14. Robert C**.** Machine learning, a probabilistic perspective. Taylor & Francis; 2014.
15. Hubel DH, Wiesel TN**.** Receptive fields and functional architecture of monkey striate cortex. The Journal of physiology. 1968;195:215-243.
16. Dutta JK, Liu J, Kurup U, et al. Effective Building Block Design for Deep Convolutional Neural Networks using Search. arXiv preprint arXiv:1801.08577. 2018.
17. Simonyan K, Zisserman A**.** Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014.
18. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. 2015. Cvpr.
19. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. 770-778.
20. Long J, Shelhamer E, Darrell T**.** Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. 3431-3440.
21. Ronneberger O, Fischer P, Brox T**.** U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention, 2015. Springer: 234-241.
22. He K, Gkioxari G, Dollár P, et al. Mask r-cnn. Computer Vision (ICCV), 2017 IEEE International Conference on, 2017. IEEE: 2980-2988.
23. Lin T-Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context. European conference on computer vision, 2014. Springer: 740-755.
24. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object

detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2014. 580-587.

25. Girshick R. Fast r-cnn. arXiv preprint arXiv:1504.08083. 2015.
26. Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 2015. 91-99.
27. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. 779-788.
28. Huang Y, Liu Z, He L, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non—Small Cell Lung Cancer. Radiology. 2016;281:947-957.
29. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. Advances in neural information processing systems, 2014. 2672-2680.
30. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434. 2015.
31. Reed S, Akata Z, Yan X, et al. Generative adversarial text to image synthesis. arXiv preprint arXiv:1605.05396. 2016.
32. Bousmalis K, Silberman N, Dohan D, et al. Unsupervised pixel-level domain adaptation with generative adversarial networks. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. 7.
33. Tzeng E, Hoffman J, Saenko K, et al. Adversarial discriminative domain adaptation. Computer Vision and Pattern Recognition (CVPR), 2017. 4.
34. Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network. arXiv preprint. 2016.
35. Sønderby CK, Caballero J, Theis L, et al. Amortised map inference for image super-resolution. arXiv preprint arXiv:1610.04490. 2016.
36. Kamel SI, Levin DC, Parker L, et al. Utilization Trends in Noncardiac Thoracic

Imaging, 2002-2014. J Am Coll Radiol. 2017;14:337-342.

37. Finigan JH, Kern JA. Lung cancer screening: past, present and future. Clin Chest Med. 2013;34:365-371.

38. Quekel LG, Kessels AG, Goei R, et al. Miss rate of lung cancer on the chest radiograph in clinical practice. Chest. 1999;115:720-724.

39. Li F, Engelmann R, Armato SG, 3rd, et al. Computer-aided nodule detection system: results in an unselected series of consecutive chest radiographs. Acad Radiol. 2015;22:475-480.

40. van Beek EJ, Mullan B, Thompson B. Evaluation of a real-time interactive pulmonary nodule analysis system on chest digital radiographic images: a prospective study. Acad Radiol. 2008;15:571-575.

41. Wang C, Elazab A, Wu J, et al. Lung nodule classification using deep feature fusion in chest radiography. Comput Med Imaging Graph. 2017;57:10-18.

42. Pesce E, Ypsilantis P-P, Withey S, et al. Learning to detect chest radiographs containing lung nodules using visual attention networks. arXiv:1712.00996. 2017.

43. Miller C, Lonnroth K, Sotgiu G, et al. The long and winding road of chest radiography for tuberculosis detection. Eur Respir J. 2017;49.

44. Melendez J, Sanchez CI, Philipsen RH, et al. An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. Sci Rep. 2016;6:25265.

45. Pande T, Cohen C, Pai M, et al. Computer-aided detection of pulmonary tuberculosis on digital chest radiographs: a systematic review. Int J Tuberc Lung Dis. 2016;20:1226-1230.

46. X. Wang YP, L. Lu, Z. Lu, M. Bagheri, R. M. Summers. ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017.

47. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68:7-30.

48. National Lung Screening Trial Research T, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365:395-409.

49. Oudkerk M, Devaraj A, Vliegenthart R, et al. European position statement on lung cancer screening. Lancet Oncol. 2017;18:e754-e766.

50. Goo JM**.** Computer-aided detection of lung nodules on chest CT: issues to be solved before clinical use. Korean J Radiol. 2005;6:62-63.

51. Armato SG, 3rd, McLennan G, Bidaut L, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. Med Phys. 2011;38:915-931.

52. Hua KL, Hsu CH, Hidayati SC, et al. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. Onco Targets Ther. 2015;8:2015-2022.

53. Setio AAA, Ciompi F, Litjens G, et al. Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks. IEEE Transactions on Medical Imaging. 2016;35:1160-1169.

54. Hamidian S, Sahiner B, Petrick N, et al. 3D Convolutional Neural Network for Automatic Detection of Lung Nodules in Chest CT. Proc SPIE Int Soc Opt Eng. 2017;10134.

55. Jiang H, Ma H, Qian W, et al. An Automatic Detection System of Lung Nodule Based on Multi-Group Patch-Based Deep Learning Network. IEEE J Biomed Health Inform. 2017.

56. Masood A, Sheng B, Li P, et al. Computer-Assisted Decision Support System in Pulmonary Cancer detection and stage classification on CT images. J Biomed Inform. 2018;79:117-128.

57. van Riel SJ, Sanchez CI, Bankier AA, et al. Observer Variability for Classification of Pulmonary Nodules on Low-Dose CT Images and Its Effect on Nodule Management. Radiology. 2015;277:863-871.

58. Ciompi F, Chung K, van Riel SJ, et al. Towards automatic pulmonary nodule

management in lung cancer screening with deep learning. Sci Rep. 2017;7:46479.

59. Nibali A, He Z, Wollersheim D**.** Pulmonary nodule classification with deep residual networks. Int J Comput Assist Radiol Surg. 2017;12:1799-1808.

60. Zhao X, Liu L, Qi S, et al. Agile convolutional neural network for pulmonary nodule classification using CT images. Int J Comput Assist Radiol Surg. 2018;13:585-595.

61. Shen W, Zhou M, Yang F, et al. Multi-crop Convolutional Neural Networks for lung nodule malignancy suspiciousness classification. Pattern Recognition. 2017;61:663-673.

62. Harari S, Caminati A, Madotto F, et al. Epidemiology, survival, incidence and prevalence of idiopathic pulmonary fibrosis in the USA and Canada. Eur Respir J. 2017;49.

63. Watadani T, Sakai F, Johkoh T, et al. Interobserver variability in the CT assessment of honeycombing in the lungs. Radiology. 2013;266:936-944.

64. Park B, Seo JB, Kim N, et al. CT Lung Image Segmentation for Diffuse Interstitial Lung Disease using Deep Convolutional Neural Networks. Medical Physics In submission.

65. Anthimopoulos M, Christodoulidis S, Ebner L, et al. Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. IEEE Trans Med Imaging. 2016;35:1207-1216.

66. Kim GB, Jung KH, Lee Y, et al. Comparison of Shallow and Deep Learning Methods on Classifying the Regional Pattern of Diffuse Lung Disease. J Digit Imaging. 2017.

67. Gao M, Bagci U, Lu L, et al. Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. Comput Methods Biomech Biomed Eng Imaging Vis. 2018;6:1-6.

68. Humphries SM, Yagihashi K, Huckleberry J, et al. Idiopathic Pulmonary Fibrosis: Data-driven Textural Analysis of Extent of Fibrosis at Baseline and 15-Month Follow-up. Radiology. 2017;285:270-278.

69. Harrison AP, Xu Z, George K, et al. Progressive and Multi-path Holistically Nested Neural Networks for Pathological Lung Segmentation from CT Images. 2017. Cham.

Springer International Publishing: 621-629.

70. Ross JC, San Jose Estepar R, Kindlmann G, et al. Automatic lung lobe segmentation using particles, thin plate splines, and maximum a posteriori estimation. Med Image Comput Comput Assist Interv. 2010;13:163-171.

71. George K, Harrison AP, Jin D, et al. Pathological Pulmonary Lobe Segmentation from CT Images Using Progressive Holistically Nested Neural Networks and Random Walker. 2017. Cham. Springer International Publishing: 195-203.

72. Park J, Seo JB, Kim N, et al. Fully automated lung lobe segmentation on volumetric chest CT with 3D U-Net; Validation with intra- and extra-datasets. Medical Physics In submission.

73. Pu J, Gu S, Liu S, et al. CT based computerized identification and analysis of human airways: a review. Med Phys. 2012;39:2603-2616.

74. Lo P, van Ginneken B, Reinhardt JM, et al. Extraction of airways from CT (EXACT'09). IEEE Trans Med Imaging. 2012;31:2093-2107.

75. Charbonnier JP, Rikxoort EMV, Setio AAA, et al. Improving airway segmentation in computed tomography using leak detection with convolutional networks. Med Image Anal. 2017;36:52-60.

76. Jin D, Xu Z, Harrison AP, et al. 3D Convolutional Neural Networks with Graph Refinement for Airway Segmentation Using Incomplete Data Labels. 2017. Cham. Springer International Publishing: 141-149.

77. Schaller S, Wildberger JE, Raupach R, et al. Spatial domain filtering for fast modification of the tradeoff between image sharpness and pixel noise in computed tomography. IEEE transactions on medical imaging. 2003;22:846-853.

78. Boedeker KL, McNitt-Gray MF, Rogers SR, et al. Emphysema: effect of reconstruction algorithm on CT imaging measures. Radiology. 2004;232:295-301.

79. Kim J, Kwon Lee J, Mu Lee K**.** Accurate image super-resolution using very deep convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 1646-1654.

80. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by

noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.

81. Lao J, Chen Y, Li ZC, et al. A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme. Sci Rep. 2017;7:10353.

82. Paul R, Hawkins SH, Schabath MB, et al. Predicting malignant nodules by fusing deep features with classical radiomics features. J Med Imaging (Bellingham). 2018;5:011021.

83. Karamitsos TD, Francis JM, Myerson S, et al. The role of cardiovascular magnetic resonance imaging in heart failure. J Am Coll Cardiol. 2009;54:1407-1424.

84. Grothues F, Smith GC, Moon JC, et al. Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-dimensional echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy. Am J Cardiol. 2002;90:29-34.

85. Sapra R, Singh B, Thatai D, et al. Critical appraisal of left ventricular function assessment by the automated border detection method on echocardiography. Is it good enough? Int J Cardiol. 1998;65:193-199.

86. Rahmouni HW, Ky B, Plappert T, et al. Clinical utility of automated assessment of left ventricular ejection fraction using artificial intelligence-assisted border detection. Am Heart J. 2008;155:562-570.

87. Winther HB, Hundt C, Schmidt B, et al. v-net: Deep learning for generalized biventricular cardiac mass and function parameters. arXiv preprint arXiv:1706.04397. 2017.

88. Tran P**.** A fully convolutional neural network for cardiac segmentation in short-axis MRI. arXiv preprint arXiv:1604.00494. 2016.

89. C. Baumgartner LK, M. Pollefeys, E. Konukoglu**.** An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. Proc. Statistical Atlases and Computational Models of the Heart (STACOM), ACDC challenge, MICCAI'17 Workshop. 2017.

90. Winther HB, Hundt C, Schmidt B, et al. nu-net: Deep Learning for Generalized

Biventricular Mass and Function Parameters Using Multicenter Cardiac MRI Data. JACC Cardiovasc Imaging. 2018.

91. Zheng Q, Delingette H, Duchateau N, et al. 3D Consistent & Robust Segmentation of Cardiac Images by Deep Learning with Spatial Propagation. IEEE. 2018.

92. Schulz-Menger J, Bluemke DA, Bremerich J, et al. Standardized image interpretation and post processing in cardiovascular magnetic resonance: Society for Cardiovascular Magnetic Resonance (SCMR) board of trustees task force on standardized post processing. J Cardiovasc Magn Reson. 2013;15:35.

93. Vovk U, Pernus F, Likar B**.** A review of methods for correction of intensity inhomogeneity in MRI. IEEE Trans Med Imaging. 2007;26:405-421.

94. Prevedello LM, Erdal BS, Ryu JL, et al. Automated Critical Test Findings Identification and Online Notification System Using Artificial Intelligence in Imaging. Radiology. 2017;285:923-931.

95. Lee SI, Krishnaraj A, Chatterji M, et al. When Does a Radiologist's Recommendation for Follow-up Result in High-Cost Imaging? Radiology. 2012;262:544-549.

Figure legends

Figure 1. Architecture of convolutional neural network (CNN): (a) A CNN is comprised of one or more convolutional layers (often with a pooling layer) and then followed by one or more fully connected layers, (b) the convolutional layers compute the sum of the element-wise multiplication between the input image and multiple filters (i.e. kernel) to detect distinctive local motifs, (c) the pooling layers, which is a form of non-linear down-sampling (e.g. max and average pooling), serve to progressively reduce the spatial size of the representation and reduce the number of parameters and amount of computation in the network; this process helps to limit overfitting of the data.

Figure 2. By a skip connection (i.e. identity mapping) and element-wise addition, a residual block makes it easier to train a deeper network without extra parameter and computational complexity.

Figure 3. Generative adversarial networks (GAN)s consist of a generator and a discriminator, wherein the generator aims to generate sample (synthetic images) that resemble those in the training data while the discriminator tries to distinguish between the two.

Figure 4. A model for detecting five kinds of pulmonary abnormalities (including nodule (ND), consolidation (CS), interstitial opacity (IO), pleural effusion (PE), and pneumothorax (PT)) on chest radiograh with weak labeled data, which indicate the presence of abnormalities' labels only. Upper rows of each case depict the regions of interest labeled by radiologists, and lower rows show the class activation map (CAM). Localization of trained abnormal patterns through CAM could assist radiologists to diagnose lung diseases much easier.

Figure 5. Detection of multiple abnormal lesions on chest radiograh. A CNN model was trained with strongly labeled data, which indicates not only the type of abnormalities but also their locations and boundaries. Multiple lesions were detected in the whole lung

images, and detected regions matched with those of interest delineated by radiologists. This approach could assist radiologists to diagnose and monitor multiple lesions of whole lung.

Figure 6. A method for detecting multi-scale nodules. By Training with RGB color images that were comprised of three adjacent slice in the axial plane, we detected nodules of various sizes. (a)-(d) depict the detected nodules with sizes of 3.4 mm, 5.6 mm, 9.8 mm, and 14.4 mm, respectively.

Figure 7. Regional image patterns of Diffuse interstitial Lung Disease (DILD) using 3D CNN. Since the diagnosis of DILD shows significant variation in inter- and intra-observer interpretation due to a lack of standard criteria and a burden of reviewing a large amount of data, CNN based automated classification on voxel-by-voxel basis is necessary for the quantification of disease extent and distribution of DILD.

Figure 8. We employed 3D U-Net (a typical type of 3D CNN) to develop a robust lobe segmentation. This approach also performed well in the fake and incomplete fissures, since this network was trained on lobe-by-lobe expert human training set.

Figure 9. The fully automated airway segmentation method in a patient with chronic obstructive lung disease (a), which started from (b) the initial airways by using the region growing method. (c) Our method achieved a high sensitivity at a low false positive rate with fast execution time (2-8 min). (d) Manual segmentation usually required 1-2 hours by an experienced research assistant.

Figure 10. Conversion of CT images reconstructed with one kernel to images with different kernels without using a sinogram: (a) CNN architecture for CT kernel conversion. (b)-(c) CT images reconstructed with B10f and B70f, respectively. (d) A CT image interconverted from B10f to B70f using (a). (e)-(f) Difference images between (b)-(c) and (b)-(d), respectively.

~~**Figure 11.** A selection of images of the Hannover Medical School (MHH), MICCAI 2009 LV~~

~~Segmentation Challenge (LVSC), and the Right Ventricular Segmentation Challenge (RVSC) data sets. From the left to the right column, the images depict the MR image, the predicted segmentation, the ground truth segmentation and the difference between ground truth and prediction, where the error is denoted in red. The upper two rows show good agreement between the predicted segmentation by v-net and the ground truth at the apex and the base. The lower three rows display the apex and base and the disagreement between the predicted segmentation and ground truth from the experts. Retrospectively, one could argue that the v-net segmentation provides a more accurate delineation of the epi- and endocardium, compared to the ground truth (with permission from v-net: Deep learning for generalized biventricular cardiac mass and function parameters. arXiv preprint arXiv:170604397 2017).~~

Commented [이상11]: R2-2

Deep Learning Applications in Chest Radiography and CT: Current State of the Art

Type of Manuscript: Invited review

Summary

Deep learning is a genre of machine learning that allows computational models to learn representations of data with multiple levels of abstraction using numerous processing layers. A distinctive feature of deep learning, compared to conventional machine learning methods, is that it can generate appropriate models for tasks directly from the raw data, removing the need for human-led feature extraction.

Medical images are particularly suited for deep learning applications. Deep learning techniques have already demonstrated high performance in detection of diabetic retinopathy on fundoscopic images and metastatic breast cancer cells on pathologic images. In radiology, deep learning has the opportunity to provide improved accuracy of image interpretation and diagnosis. Many groups are exploring the possibility of using deep learning based applications to solve unmet clinical needs.

In chest imaging, there has been a large effort to develop and apply computer-aided detection (CAD) systems for the detection of lung nodules on chest radiographs and chest computed tomography. The essential limitation to CAD is an inability to learn from new information. To overcome these deficiencies, many groups have turned to deep learning approaches with promising results. In addition to nodule detection, interstitial lung disease recognition, lesion segmentation, diagnosis and patient outcomes have been addressed by deep learning approaches.

The purpose of this review article is to cover the current state of the art for deep learning approaches and its limitations, and some of the potential impact on the field of radiology, with specific reference to chest imaging.

Keywords

Chest imaging; Machine learning; Deep learning; Radiography; Computed tomography; Magnetic resonance imaging

Introduction

Deep learning is a genre of machine learning that allows computational models to learn representations of data with multiple levels of abstraction through the use of a number of unique processing layers [1]. The most distinctive feature of deep learning, compared to the conventional machine learning methods, is that it can extract fully automated features and generate appropriate models for tasks directly from the raw data on its own, removing the need for human-led feature extraction.

In recent years, deep learning methods have shown breakthroughs in various fields including image recognition [2], speech recognition [3] as well as information technology. However, in the medical field, the application of deep learning is currently in its infancy. Medical images and their respective patient electronic medical records are well suited for analysis by deep learning. Some of the first successful demonstrations of deep learning techniques were reported in the detection of lymph node metastasis from hematoxylin and eosin stained pathologic micrographs, analysis of skin cancer from photographs of the lesion, and the diagnosis of diabetic retinopathy from fundoscopic images [4-7].

In radiology, deep learning will help to improve efficiency by automated image interpretation and generation of an appropriate differential diagnosis. Data mining of the patient's electronic medical record data (big data) combined with deep learning applied to the patient's medical images should help to improve patient outcomes. Cloud based applications also allow the deep learning algorithm to continuously learn on data sets that are not restricted to a single institution. Many groups are now exploring deep learning-based applications for solutions to unmet clinical needs. In chest imaging, significant effort has been directed at developing and applying computer-aided detection (CAD) [8] systems for the detection of nodules on chest radiographs and chest computed tomography (CT) [9, 10]. Although many CAD systems are being used in clinical practice, the implementation of CAD has not been widely accepted due to its poor performance (i.e. frequent false positive and false negative cases). Deep learning approaches have the potential to overcome the limitations of existing CAD systems, with several studies showing promising results [11, 12]. Moreover, disease pattern recognition, lesion segmentation, diagnosis and survival prediction have been successfully studied using deep learning in chest imaging [13]. However, there is still concern on this technology in terms of clinical application.

In this review article, we introduce the principle methods of deep learning, their potential applications and clinical promise in chest imaging.

Deep Learning and Convolutional Neural Networks

Machine learning is defined as a set of methods that can automatically detect patterns in data, and then utilize the uncovered patterns to classify, predict, or conduct various types of decision making under uncertain conditions [14]. Conventional machine learning techniques rely on extensive data engineering and considerable domain expertise to design a "feature extractor" algorithm that converts the raw data into suitable representations for computational analysis. A convolutional neural network (CNN) is a special type of deep learning and is quite similar to the overall learning process (e.g. neuronal pruning) of the mammalian visual cortex [15], and is responsible for the recent improvements in the field of computer vision (e.g. self-driving automobiles). With the availability of large datasets and increased computing power, CNNs have produced promising results for many tasks including image classification, correct image detection, correct image segmentation and understanding speech (e.g. natural language processing).

The architecture of a CNN can be composed of convolutional, pooling, and fully connected layers (Figure 1): (1) the convolutional layers detect distinctive local motifs by applying multiple filters and generating multiple feature maps, (2) the pooling layers effectively reduce the dimensions of feature maps (other techniques such as dilated convolutions, convolutions with stride can also be used), (3) the fully connected layers integrate all feature responses and eventually project onto an output layer which serves to answer the task at hand. By using deep CNN architecture (repeating the convolutional and pooling layers several times) to mimic the natural neuromorphic multi-layer network, deep learning can automatically and adaptively learn a hierarchical representation of patterns and consequently identify the most significant features for a given task [2]. In order to deal with complex tasks, networks with many layers – so called Deep Networks – are required. However, adding additional layers increases the number of parameters in the model and can make it more difficult to train it for a specific task without overfitting the data.

1. Classification

One key task for radiologists is creating an appropriate differential diagnosis for each patient's medical images. This job can be computationally defined as a typical classification task using input from medical images and any available clinical information. There are many different CNN network architectures for classifying images. In order to improve the efficiency of the training procedure and reduce the number of parameters, the deeper networks have introduced more effective subroutines - "building blocks". These building blocks are small branching/spanning convolution blocks with pooling and batch normalization layers, which can be repeated to construct deeper architectures [16]. VGG19 [17] used small and fixed size kernels in each layer to win the ImageNet challenge of 2014. Another CNN named GoogLeNet (i.e. " Inception") [18] made use of the building block that is a multi-level feature extractor with a set of convolutions of different sizes. ResNet [19], which won the ImageNet challenge of 2015, introduced the subroutine of a "residual building block" which was designed to learn the residual (e.g. features that remain important) in order to make it easier to train deeper neural networks. This residual block was implemented by adding the input of the block to the output of the layers within the block (Figure 2). Since 2014, the performance on the ImageNet benchmark has saturated, but use of these architectures remains popular for medical image processing.

Moreover, these days, various combinations of these building blocks are used to construct deep learning architectures for the desired tasks, instead of relying on the bare previously proposed architecture.

2. Segmentation

The innovations of object classification have now shifted to semantic segmentation. This is a common task for both natural and medical image analysis whereby each voxel is classified in an image to determine the boundary conditions that define a specific object. The fully convolutional network (FCN) [20] represents a critical breakthrough for deep learning based semantic segmentation. In a FCN, the fully connected layers in the standard CNNs are replaced by convolutions with large receptive fields. This method achieves this segmentation by using coarse class score maps obtained by feed forwarding an input image. U-Net [21], which is the most well-known segmentation architecture in medical image analysis, combines an equal amount of upsampling and downsampling layers with skip connections between opposing convolution and deconvolution layers. Mask Region-based CNN (R-CNN) [22] detects objects in an image while simultaneously generating segmentation masks for each instance. This method has achieved the state-of-the-art performance on Microsoft Common Objects in Context [23].

3. Detection

The detection of objects of interest (i.e. lesions) is a key part of diagnosis and is one of the most labor-intensive tasks for radiologists. Several CNN network architectures have been shown to be able to detect a variety of objects quickly and accurately. R-CNN [24] combines region proposals (from a defined set of candidate detections) with CNNs and has then been improved to Fast R-CNN [25] and Faster R-CNN [26] with better performance. There are a few methods that computationally approach the problem of image detection by using multivariate regression; two of the most popular CNNs are You only look once (YOLO) [27] and Single shot multibox detector (SSD) [28], successfully predicting bounding boxes and classification probability.

4. Generative Model

Generative adversarial networks (GANs) have the advantage of automatically producing new images (e.g. synthetic image data) similar to samples from the training set by using two competing CNNs where one is generating artificial samples and the other is discriminating artificial from real samples [29] (Figure 3). These GANs could be trained end-to-end and learn representative features in a completely unsupervised manner. The representations learned by GANs are employed in various applications including medical image syntheses [30, 31], image normalization [32, 33], and super-resolution [34, 35].

Applications in Chest Imaging

I. Chest Radiography

Chest radiography is the most commonly performed diagnostic imaging procedure and over 35 million chest radiographs are performed each year in the United States alone and the average radiologist reads more than 100 chest radiograph exams per day [36]. Although these exams are clinically useful, efficient, and cost-effective, chest radiography consists of complex 3-dimensional anatomic information condensed in a 2-dimensional projection. Accurate interpretation of chest radiographs requires a great deal of experience and medical knowledge on the part of the radiologist. Increased radiologist work-loads combined with the intrinsic challenges of interpreting chest radiographs is associated with considerable inter- and intra-reader variability, missed lesions, and reporting delays in today's medical practice [8]. Deep learning technology has the potential to automatically detect abnormalities or assist radiologists in reading chest radiographs. Such technology would be very attractive for rural areas with few radiologists as well as for state-of-the-art medical centers to help support high volume workflows and improve efficiency of the radiology departments [12].

1. Lung Nodule Detection

Lung nodule detection from chest radiographs is another promising area for the application of deep learning technology. Lung cancer is the leading cause of cancer death worldwide and chest radiography has been the most widely adopted screening and imaging tool to detect lung cancer. However, unfortunately, due to the confounding effects of anatomic complexity on chest radiographs, lung cancer screening using plain chest film has yielded unsatisfying results, with reports of missed nodules being as high as 40% [37, 38].

CAD systems have been developed to help radiologists detect lung nodules. Recently this method has shown a sensitivity of 71% with 1.3 false-positive CAD marks per image[39]. Using bone-suppressed dual energy chest radiographs, another stand-alone

CAD system achieved a sensitivity of 74% with a 1.0 false-positive CAD mark per image [9]. In the setting of follow-up of patients with previous cancer (of any type), another CAD system showed promise by improving sensitivity from 63% to 92% for detection of lung nodules while only slightly decreasing specificity (from 98% to 96%) [40]. Although there has been improvement in CAD nodule detection on chest radiograph, these methods still need better accuracy before they are routinely accepted.

Recently CAD systems using deep learning techniques have shown improved accuracy for nodule detection on chest radiograph. A deep learning based technique identified by Wang et al. extracted deep learning features by transfer learning and combined them with traditional hand-crafted features. This CAD system achieved a higher sensitivity (69.3%) for nodule detection at a significantly lower false positive rate (1.19 false positive marks per exam) [41]. There is a recent report using CNN with visual attention networks generating respective accuracies of 0.76 for nodule detection and 0.65 for nodule localization on chest radiographs [42].

2. Diagnosis of Tuberculosis

Another specific field of research with great potential benefit to public health is utilizing deep learning technology for the diagnosis of pulmonary mycobacterium tuberculosis (TB) based on chest radiography. TB is an infectious disease caused by the acid fast (i.e. outer capsule appears red on hematoxylin and eosin staining) bacillus mycobacterium tuberculosis. In western countries, it is often thought as a "disease of the past", but TB is still a major health problem in the developing world, with millions of new cases encountered every year.

While the diagnosis of TB can be confirmed by bacteriology or the whole blood gamma interferon release assay (Quantiferon Gold, Qiagen), chest radiography is a highly sensitive imaging tool for triaging and screening for current and previous pulmonary TB infection. This organism is very difficult to culture in vitro and is often not able to be confirmed bacteriologically [43]. In locations where the prevalence of TB is high, there are a limited number of experienced chest radiologists available to use chest radiography as a method to confirm this disease. This shortage impairs screening efficacy and limits the opportunity to start medical therapy for a complete recovery [44]. Therefore, considerable effort developing CAD systems for the detection of pulmonary TB on chest radiographs has been extended. Traditional CAD systems without deep learning technology have shown acceptable TB detection performance with an area under the curve ranging from 0.71 to 0.84 [45].

Recently Lakhani et al. reported the performance of CAD using CNN for detection of pulmonary TB [12] and in that study, the CAD system reached an AUC of 0.99, which is greater than any previously reported CAD system. Although external validation is still needed to determine the true clinical benefit, CNN based CAD is a feasible and promising approach in this clinical scenario.

3. Multiple Abnormal Pattern detection

Although the detection of lung nodules and TB on chest radiographs has gained attention for many deep learning researchers, these findings can be relatively rare. Each chest radiograph may contain many abnormalities, for example: pneumonia, pleural effusions, pneumothorax, medical devices and cardiomegaly (Figure 4, 5). Therefore, facilitating deep learning technology to detect multiple abnormal patterns (MAP), rather than concentrating simply on nodules or TB would be more clinically practical.

Emergence of deep learning has drastically improved the performance of machine learning for object recognition, detection, and localization when compared to previous methodologies. Critical to the success of these methods are well-annotated ("strongly labeled") large datasets for effective system training. Recently two large datasets of chest radiographs, Open-I and ChestX-ray14, consisting of more than 110,000 chest radiographs from 30,805 patients, have been publicly released and have attracted considerable attention in the deep learning community. These publicly available data are external validation sets for any deep learning application using chest radiographs

In 2017, Wang et al. trained various known CNN models to detect 8 abnormal patterns (atelectasis, cardiomegaly, effusion, infiltration, mass, nodules, pneumonia, pneumothorax) on chest radiographs and achieved accuracy ranging from 0.56 to 0.78 [46]. Another study by Cicero et al. reported that a retrospective analysis of 35,038 chest radiographs from a single medical center using CNN (GoogLeNet), they were able to obtain MAP classification accuracy of 0.88 [11]. MAP detection in CXR with deep learning technology is still an area of active ongoing research and different methodologies are being tested and validated, and overall accuracy will likely improve.

II. Chest Computed Tomography

Unlike chest radiographs, chest CT provides cross-sectional images, allowing for direct 3-dimensional visualization of anatomic structures. Chest CT has a much higher sensitivity and lower inter-reader variability for detection of lung abnormalities and is frequently employed in the diagnosis and follow-up of most pulmonary diseases. Additionally, enhanced clinical availability, decreased cost, reduced radiation dose, and overall technical improvements of CT machines have resulted in a progressive increase in numbers of CT exams performed each year. Therefore, effective CAD system for chest CT interpretation would promote overall workflow for radiologists, by reducing the time required to read each CT exam and enhance reading accuracy.

1. Nodule Detection / Screening

Accurate nodule detection on chest CT has become a recent point of emphasis for efficient lung cancer screening. Despite advances in cancer treatment and screening programs, most lung cancer patients are still initially diagnosed at an advanced stage of the disease, which is associated with less than 20% 5-year survival [47].

Since the National Lung Screening Trial (NLST) announced a significant improvement (20%) in lung cancer mortality in high-risk populations when screened with low-dose chest CT (LDCT) [48], LDCT for cancer screening has been widely accepted [49]. Potentially this will lead to an increased volume of LDCT, which will require expert analysis from a radiologist for the detection and classification of nodules into either benign or malignant diagnoses.

A CAD system could aid radiologists in both detection and classification of lung nodules (Figure 6). Although traditional CAD systems have provided solid results, they often consist of complex pipelines of algorithms that depend heavily on manual human input such as pre-processing, segmentation, feature extraction, and model training, potentially hindering their performance [50]. Application of deep learning technology, on the other hand, can potentially remove innate challenges in traditional CAD systems by providing seamless feature identification and classification and removing the need for complex human-led feature extraction pipelines.

In 2011, the Lung Image Database Consortium (LIDC) database, containing 1018 cases of thoracic CT scans and image annotations by 4 thoracic radiologists, was released and has motivated deep learning researchers to develop CAD systems for chest CT nodule detection and classification [51]. CNNs are the most commonly utilized deep learning technology for lung nodule detection on CT images, and achieves good nodule detection sensitivity while maintaining an acceptable false positive rate. The first report of CAD system with deep learning technology for lung nodule detection on CT was Hua et al. in 2015, achieving sensitivity of 73% and specificity of 80%, which was superior to any other available conventional CAD systems [52]. In 2016, Setio et al. trained CNN to detect pulmonary nodules and achieved 85.4% sensitivity with only one false positive lesion per scan [53]. Studies that are more recent have shown the ability of CNNs to boost nodule detection sensitivity on CT to a higher level (95%) but were associated with a wide range (1.17 - 22.4) of false positive rates [54-56].

Classification of detected lung nodules is also a potential area that could benefit from the use of CAD systems. CT characteristics of a lung nodule, mainly nodule type and size, are closely associated with the likelihood of malignancy. These CT features are important determinants for planning treatment and follow up strategy. However, there is considerable observer variability in classification of pulmonary nodules among radiologists and this can lead to redundant follow-up examinations, unnecessary invasive procedures, or neglected malignancy [57]. In 2017, Ciompi et al. introduced a deep learning system that achieved good performance for nodule type classification based on lung-RADS system and was even within the inter-observer variability among four experienced human readers [58]. Furthermore, one study found that nodule classification accuracy of CAD system was improved by combining deep residual learning, curriculum learning, and transfer learning [59]. Other studies using different CNN models have achieved a classification accuracy as high as 87.1% [60, 61].

2. Interstitial Lung Disease

Interstitial lung disease (ILD) pattern classification is another area of research for deep learning technology. ILD is characterized by progressive fibrosis or inflammation of lung tissue and eventual deterioration of respiratory function [62]. Accurate diagnosis of ILD presents a challenge for the multidisciplinary medical panel at each institution that cares for these disorders because most ILDs have similar clinical manifestations, despite the fact that they are a histologically heterogeneous group of diseases with distinct prognoses.

High-resolution CT (HRCT) is currently the diagnostic imaging tool of choice for the diagnosis and evaluation of ILDs. However, ILDs have similar appearance on CT and CT readings are prone to high inter- and intra-observer variability [63]. Therefore, automatic identification and classification of different ILD patterns on chest CT may be helpful even for experienced chest radiologists, and application of deep learning technology could play an eminent role in developing such CAD systems. Segmentation of the lung with ILD could be enhanced by semantic segmentation with CNN [64]. In 2016, deep learning technology with CNN showed accuracy of 85% for classifying 6 different ILD patterns in dataset of 14,696 image patches [65]. In 2017, Kim et al. compared shallow and deep learning methods on classifying six ILD patterns on CT and found that deep learning methods showed significantly better accuracy and that accuracy was further increased with addition of more convolution layers [66] (Figure 7). More recently, a new CNN method achieved an ILD pattern classification accuracy of 87.9% using the holistic input of the entire CT data set [67]. Moreover, CAD methodology demonstrated a prognostic ability of lung function decline using quantifiable ILD on CT studies [68].

3. Chronic Obstructive Pulmonary Disease

A more basic field of application for deep learning technology is the segmentation and reconstruction of organs-of-interest from chest CT scans. Organ segmentation usually is the first step of many CAD systems, even those using deep learning methods, and accuracy of segmentation process is critical because any errors in this process would affect all the subsequent analysis. Various methods for organ segmentation have been developed and tested, showing promising results, but deep-learning based models could potentially improve methodological robustness and generalizability across imaging platforms, thus providing outcomes that are more reliable.

In 2017, Harrison et al. developed a deep model called progressive holistically-nested networks (P-HNNs) and reported that their P-HNNs model showed significant improvements of lung segmentation performance compared to previous segmentation approaches [69]. As for lobar segmentation, traditional methods are semi-automatic at best and largely relied on airway or vessel anatomy to delineate the lobar borders, with only few exceptions [70]. To address these problems, a deep learning method for lobe segmentation was introduced in 2017 and this method achieved high accuracy without reliance on prior airway or vessel segmentations, even when tested in lungs that had underlying disease [71, 72] (Figure 8).

Aside from lung tissue segmentation, robust and reliable airway segmentation is also essential for quantitative evaluation of various diseases involving the airways, such as chronic obstructive pulmonary disease (Figure 9). A large number of prior methods have common limitations that they are substantially influenced by morphologic changes in airway trees and measurement errors, such as airway leak that are most prevalent at smaller (or more peripheral) airways [73]. In fact, 15 different traditional algorithms were evaluated at an airway segmentation challenge in 2009 (EXACT 09), and precise delineation of a small bronchus without airway measurement leaks remained a common unsolved problem from this challenge [74]. In 2017, a deep learning method was developed and tested using a dataset from EXACT 09, and found that CNN significantly decreased airway leaks during segmentation process, resulting in higher sensitivity and specificity compared to all the other algorithms that participated in the EXACT 09 challenge [75]. In another study, even with incompletely annotated data, 3D deep fully convolutional networks demonstrated considerable improvements in airway segmentation while maintaining acceptable quantity of airway leaks [76].

4. Image Normalization

The reconstruction kernel is one of the most important technical parameters that determine the trade-off between spatial resolution and image noise in CT [77]. Since the selection of kernel affects quantitative analysis [78], CT images with different reconstruction kernels are necessary for various diagnostic or quantitative purposes. To overcome the limitation that it is difficult to save the raw data before reconstruction with various kernels, post-processing techniques have been developed to permit interconversion among CT images obtained with different kernels. Kim et al. [79] recently demonstrated that CNNs could be taught differences between high- and low-resolution images (residual images) and then they could be used to accurately and rapidly convert low-resolution images to high-resolution images. This approach is also applicable for interconverting CT images obtained using different kernels (Figure 10).

5. Radiomics and Deep Survival

Radiomics and prediction of patient outcomes (a.k.a." deep survival") are also active areas of research for the application of deep learning technology. Radiomics, which has gained substantial interest from researchers around the globe, involves the high-throughput extraction of quantitative features from medical images to develop reliable models to predict genomic information, clinical outcomes, and survival [80]. Extraction of radiomics features is a critical process in radiomics research and the majority of previous studies use handcrafted features, which are limited by current medical knowledge and human observation. On the other hand, CNN and transfer learning can be incorporated into radiomics models to extract more diverse features (deep features), which are free from prerequisite medical knowledge and biases. In this context, Lao et al. extracted 98,304 deep features (this would qualify as an example of over-fitting of the data) from images of glioblastoma multiforme and found 6 deep features that could predict overall survival with a concordance index of 0.71 [81].

In chest imaging, Paul et al. combined deep features of lung nodules detected on chest CT with traditional radiomics features to predict the probability of a malignant nodule and reported an overall accuracy of 76.8% and an AUC of 0.87 [82]. Another group used CNN to predict patient outcomes in a large cohorts of smokers and COPD patients and the CNN model predicted mortality with fair discrimination [13]. Deep radiomics and deep survival are promising new fields for study.

Perspective, Challenges, and Limitations

In this manuscript, we reviewed basic concepts of deep learning and its various applications in chest radiography and CT. In comparison to CT, MRI is more challenging for deep learning application because there is no pulse sequence dependent standardized intensity scale like the Hounsfield units in CT [83]. The application of this new technology to radiology has barely started, but it has shown remarkable results when compared with previous studies. We believe that these improvements in performance will soon offer new possibilities for the clinical practice of radiology.

The first deep learning based CAD application may be used to find critical findings on chest radiograph and triage the worklist before a radiologist's read. In brain CT, Prevedello et al. [84] already demonstrated that deep learning based algorithm could automatically identify critical findings and notify the interpreting radiologist. Furthermore, if the performance of CAD can be clinically acceptable in terms of prioritization of chest radiograph, it implies that deep learning based CAD have potential to differentiate normal chest radiographs from grossly abnormal exams. Thus, deep learning based CAD should improve the workflow and efficiency of radiology departments.

Second, CAD can help diagnosis of disease such as ILD and generate a preliminary quantitative report based on CAD results. This CAD report is repeatable with the same results and has no "intra-reader" variability. CAD combined with big data technology may retrieve similar images or diagnosis when radiologists require during interpretation of CT. It can also help to reduce the reading time.

Third, automation of lesion detection, segmentation, quantification by deep learning techniques facilitates reporting of the quantitative analysis of medical images more easily. Deep learning based segmentation tool improves accuracy and decreases image interpretation time. Furthermore, these data will likely provide improve the prediction of patient outcomes and risk stratification.

However, there are still many challenges to overcome. Currently, training deep learning algorithms requires large, strongly labeled and anonymized image datasets. These data sets are very challenging to acquire. While some abnormalities such as pneumothorax and malpositioned lines/tubes can be based upon imaging findings alone, most diseases require clinical documentation and/or pathological confirmation. Ambiguous or overlapping radiographic terms such as "consolidation" and "infiltrate" should not be used as surrogates for pneumonia to label training cases. This has been recognized as a limitation of some publically available datasets. National organizations (e.g. ACR, RSNA) and subspecialty radiology societies can play important roles in defining appropriate tasks for deep learning algorithms, as well as assisting in making publicly available strongly labeled training data sets and validation data sets.

Futhermore, the challenges regarding the ethical and legal aspects of data sharing and patient privacy are also paramount. There are severe monetary penalties (i.e. fines) in the United States of America [85] for any medical facility that allows compromise of personal health information/images. In the USA, the Health Insurance Portability and Accountability Act (HIPPA) governs any use of a patient's health information; as such, it is of critical importance that these imaging and medical data that are used for training, testing and validation of deep learning methods are fully anonymized and comply with this law. New data protection laws have also been introduced throughout Europe. As deep learning requires an enormous amount of high-quality data, the laws governing the safe handling of medical images and medical record data need to be followed. New technology, such as Blockchain, may be helpful in guarantying secure data sharing.

Lastly, we should demand a thorough and systematic clinical validation of any deep learning based applications as a prerequisite to commercial application. A well-known problem with these methods is overfitting and lack of utility when asked to analyze other data sets (e.g. poor interoperability). Most machine learning publications have shown their results in carefully preselected and enriched test sets (e.g. spiked to favor that algorithm with a higher prevalence of the condition than is found clinically). Thus, beyond just determining the feasibility of using any deep learning application in a test set chosen by the author, each deep learning application should be tested by a publicly available external validation set. We believe that this should be a requirement for any commercially approved deep learning method.

Conclusion

The application of deep learning methodology to help solving many tasks associated with medical imaging is at its infancy. While there are problems with every disruptive technological innovation, we believe that deep learning will soon be an indispensable tool for radiology. This is analogous to how the picture archiving communication systems and radiology information systems have transformed medical imaging and improved radiology while at the same time decreasing the cost of medical care. Reasonable expectations for this disruptive technology are needed, along with careful attention to any ethical, legal, and regulatory issues that may arise. This technology will enable radiologists to become more productive and improve patient care. The full potential of this technology will require radiologists to have an active role in governing its successful introduction to the clinic.

References

1. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521:436-444.
2. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012. 1097-1105.
3. Hinton G, Deng L, Yu D, et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. IEEE Signal Processing Magazine. 2012;29:82-97.
4. Gulshan V, Peng L, Coram M, et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA. 2016;316:2402-2410.
5. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017;542:115-118.
6. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA. 2017;318:2199-2210.
7. Ting DSW, Cheung CY, Lim G, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. JAMA. 2017;318:2211-2223.
8. McAdams HP, Samei E, Dobbins J, 3rd, et al. Recent advances in chest radiography. Radiology. 2006;241:663-683.
9. Schalekamp S, van Ginneken B, Koedam E, et al. Computer-aided detection improves detection of pulmonary nodules in chest radiographs beyond the support by bone-suppressed images. Radiology. 2014;272:252-261.
10. Liang M, Tang W, Xu DM, et al. Low-Dose CT Screening for Lung Cancer: Computer-aided Detection of Missed Lung Cancers. Radiology. 2016;281:279-288.
11. Cicero M, Bilbily A, Colak E, et al. Training and Validating a Deep Convolutional Neural Network for Computer-Aided Detection and Classification of Abnormalities on Frontal Chest Radiographs. Invest Radiol. 2017;52:281-287.

12. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology. 2017;284:574-582.
13. Gonzalez G, Ash SY, Vegas-Sanchez-Ferrero G, et al. Disease Staging and Prognosis in Smokers Using Deep Learning in Chest Computed Tomography. Am J Respir Crit Care Med. 2018;197:193-203.
14. Robert C. Machine learning, a probabilistic perspective. Taylor & Francis; 2014.
15. Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex. The Journal of physiology. 1968;195:215-243.
16. Dutta JK, Liu J, Kurup U, et al. Effective Building Block Design for Deep Convolutional Neural Networks using Search. arXiv preprint arXiv:1801.08577. 2018.
17. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014.
18. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. 2015. Cvpr.
19. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. 770-778.
20. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. 3431-3440.
21. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention, 2015. Springer: 234-241.
22. He K, Gkioxari G, Dollár P, et al. Mask r-cnn. Computer Vision (ICCV), 2017 IEEE International Conference on, 2017. IEEE: 2980-2988.
23. Lin T-Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context. European conference on computer vision, 2014. Springer: 740-755.
24. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object

detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2014. 580-587.

25. Girshick R. Fast r-cnn. arXiv preprint arXiv:1504.08083. 2015.

26. Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 2015. 91-99.

27. Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. 779-788.

28. Huang Y, Liu Z, He L, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non—Small Cell Lung Cancer. Radiology. 2016;281:947-957.

29. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. Advances in neural information processing systems, 2014. 2672-2680.

30. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434. 2015.

31. Reed S, Akata Z, Yan X, et al. Generative adversarial text to image synthesis. arXiv preprint arXiv:1605.05396. 2016.

32. Bousmalis K, Silberman N, Dohan D, et al. Unsupervised pixel-level domain adaptation with generative adversarial networks. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. 7.

33. Tzeng E, Hoffman J, Saenko K, et al. Adversarial discriminative domain adaptation. Computer Vision and Pattern Recognition (CVPR), 2017. 4.

34. Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network. arXiv preprint. 2016.

35. Sønderby CK, Caballero J, Theis L, et al. Amortised map inference for image super-resolution. arXiv preprint arXiv:1610.04490. 2016.

36. Kamel SI, Levin DC, Parker L, et al. Utilization Trends in Noncardiac Thoracic

Imaging, 2002-2014. J Am Coll Radiol. 2017;14:337-342.

37. Finigan JH, Kern JA. Lung cancer screening: past, present and future. Clin Chest Med. 2013;34:365-371.

38. Quekel LG, Kessels AG, Goei R, et al. Miss rate of lung cancer on the chest radiograph in clinical practice. Chest. 1999;115:720-724.

39. Li F, Engelmann R, Armato SG, 3rd, et al. Computer-aided nodule detection system: results in an unselected series of consecutive chest radiographs. Acad Radiol. 2015;22:475-480.

40. van Beek EJ, Mullan B, Thompson B. Evaluation of a real-time interactive pulmonary nodule analysis system on chest digital radiographic images: a prospective study. Acad Radiol. 2008;15:571-575.

41. Wang C, Elazab A, Wu J, et al. Lung nodule classification using deep feature fusion in chest radiography. Comput Med Imaging Graph. 2017;57:10-18.

42. Pesce E, Ypsilantis P-P, Withey S, et al. Learning to detect chest radiographs containing lung nodules using visual attention networks. arXiv:1712.00996. 2017.

43. Miller C, Lonnroth K, Sotgiu G, et al. The long and winding road of chest radiography for tuberculosis detection. Eur Respir J. 2017;49.

44. Melendez J, Sanchez CI, Philipsen RH, et al. An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. Sci Rep. 2016;6:25265.

45. Pande T, Cohen C, Pai M, et al. Computer-aided detection of pulmonary tuberculosis on digital chest radiographs: a systematic review. Int J Tuberc Lung Dis. 2016;20:1226-1230.

46. X. Wang YP, L. Lu, Z. Lu, M. Bagheri, R. M. Summers. ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017.

47. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68:7-30.

48. National Lung Screening Trial Research T, Aberle DR, Adams AM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365:395-409.

49. Oudkerk M, Devaraj A, Vliegenthart R, et al. European position statement on lung cancer screening. Lancet Oncol. 2017;18:e754-e766.

50. Goo JM. Computer-aided detection of lung nodules on chest CT: issues to be solved before clinical use. Korean J Radiol. 2005;6:62-63.

51. Armato SG, 3rd, McLennan G, Bidaut L, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. Med Phys. 2011;38:915-931.

52. Hua KL, Hsu CH, Hidayati SC, et al. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. Onco Targets Ther. 2015;8:2015-2022.

53. Setio AAA, Ciompi F, Litjens G, et al. Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks. IEEE Transactions on Medical Imaging. 2016;35:1160-1169.

54. Hamidian S, Sahiner B, Petrick N, et al. 3D Convolutional Neural Network for Automatic Detection of Lung Nodules in Chest CT. Proc SPIE Int Soc Opt Eng. 2017;10134.

55. Jiang H, Ma H, Qian W, et al. An Automatic Detection System of Lung Nodule Based on Multi-Group Patch-Based Deep Learning Network. IEEE J Biomed Health Inform. 2017.

56. Masood A, Sheng B, Li P, et al. Computer-Assisted Decision Support System in Pulmonary Cancer detection and stage classification on CT images. J Biomed Inform. 2018;79:117-128.

57. van Riel SJ, Sanchez CI, Bankier AA, et al. Observer Variability for Classification of Pulmonary Nodules on Low-Dose CT Images and Its Effect on Nodule Management. Radiology. 2015;277:863-871.

58. Ciompi F, Chung K, van Riel SJ, et al. Towards automatic pulmonary nodule

management in lung cancer screening with deep learning. Sci Rep. 2017;7:46479.

59. Nibali A, He Z, Wollersheim D. Pulmonary nodule classification with deep residual networks. Int J Comput Assist Radiol Surg. 2017;12:1799-1808.

60. Zhao X, Liu L, Qi S, et al. Agile convolutional neural network for pulmonary nodule classification using CT images. Int J Comput Assist Radiol Surg. 2018;13:585-595.

61. Shen W, Zhou M, Yang F, et al. Multi-crop Convolutional Neural Networks for lung nodule malignancy suspiciousness classification. Pattern Recognition. 2017;61:663-673.

62. Harari S, Caminati A, Madotto F, et al. Epidemiology, survival, incidence and prevalence of idiopathic pulmonary fibrosis in the USA and Canada. Eur Respir J. 2017;49.

63. Watadani T, Sakai F, Johkoh T, et al. Interobserver variability in the CT assessment of honeycombing in the lungs. Radiology. 2013;266:936-944.

64. Park B, Seo JB, Kim N, et al. CT Lung Image Segmentation for Diffuse Interstitial Lung Disease using Deep Convolutional Neural Networks. Medical Physics In submission.

65. Anthimopoulos M, Christodoulidis S, Ebner L, et al. Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. IEEE Trans Med Imaging. 2016;35:1207-1216.

66. Kim GB, Jung KH, Lee Y, et al. Comparison of Shallow and Deep Learning Methods on Classifying the Regional Pattern of Diffuse Lung Disease. J Digit Imaging. 2017.

67. Gao M, Bagci U, Lu L, et al. Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks. Comput Methods Biomech Biomed Eng Imaging Vis. 2018;6:1-6.

68. Humphries SM, Yagihashi K, Huckleberry J, et al. Idiopathic Pulmonary Fibrosis: Data-driven Textural Analysis of Extent of Fibrosis at Baseline and 15-Month Follow-up. Radiology. 2017;285:270-278.

69. Harrison AP, Xu Z, George K, et al. Progressive and Multi-path Holistically Nested Neural Networks for Pathological Lung Segmentation from CT Images. 2017. Cham.

Springer International Publishing: 621-629.

70. Ross JC, San Jose Estepar R, Kindlmann G, et al. Automatic lung lobe segmentation using particles, thin plate splines, and maximum a posteriori estimation. Med Image Comput Comput Assist Interv. 2010;13:163-171.

71. George K, Harrison AP, Jin D, et al. Pathological Pulmonary Lobe Segmentation from CT Images Using Progressive Holistically Nested Neural Networks and Random Walker. 2017. Cham. Springer International Publishing: 195-203.

72. Park J, Seo JB, Kim N, et al. Fully automated lung lobe segmentation on volumetric chest CT with 3D U-Net; Validation with intra- and extra-datasets. Medical Physics In submission.

73. Pu J, Gu S, Liu S, et al. CT based computerized identification and analysis of human airways: a review. Med Phys. 2012;39:2603-2616.

74. Lo P, van Ginneken B, Reinhardt JM, et al. Extraction of airways from CT (EXACT'09). IEEE Trans Med Imaging. 2012;31:2093-2107.

75. Charbonnier JP, Rikxoort EMV, Setio AAA, et al. Improving airway segmentation in computed tomography using leak detection with convolutional networks. Med Image Anal. 2017;36:52-60.

76. Jin D, Xu Z, Harrison AP, et al. 3D Convolutional Neural Networks with Graph Refinement for Airway Segmentation Using Incomplete Data Labels. 2017. Cham. Springer International Publishing: 141-149.

77. Schaller S, Wildberger JE, Raupach R, et al. Spatial domain filtering for fast modification of the tradeoff between image sharpness and pixel noise in computed tomography. IEEE transactions on medical imaging. 2003;22:846-853.

78. Boedeker KL, McNitt-Gray MF, Rogers SR, et al. Emphysema: effect of reconstruction algorithm on CT imaging measures. Radiology. 2004;232:295-301.

79. Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 1646-1654.

80. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by

noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.

81. Lao J, Chen Y, Li ZC, et al. A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme. Sci Rep. 2017;7:10353.

82. Paul R, Hawkins SH, Schabath MB, et al. Predicting malignant nodules by fusing deep features with classical radiomics features. J Med Imaging (Bellingham). 2018;5:011021.

83. Vovk U, Pernus F, Likar B**.** A review of methods for correction of intensity inhomogeneity in MRI. IEEE Trans Med Imaging. 2007;26:405-421.

84. Prevedello LM, Erdal BS, Ryu JL, et al. Automated Critical Test Findings Identification and Online Notification System Using Artificial Intelligence in Imaging. Radiology. 2017;285:923-931.

85. Lee SI, Krishnaraj A, Chatterji M, et al. When Does a Radiologist's Recommendation for Follow-up Result in High-Cost Imaging? Radiology. 2012;262:544-549.

Figure legends

Figure 1. Architecture of convolutional neural network (CNN): (a) A CNN is comprised of one or more convolutional layers (often with a pooling layer) and then followed by one or more fully connected layers, (b) the convolutional layers compute the sum of the element-wise multiplication between the input image and multiple filters (i.e. kernel) to detect distinctive local motifs, (c) the pooling layers, which is a form of non-linear down-sampling (e.g. max and average pooling), serve to progressively reduce the spatial size of the representation and reduce the number of parameters and amount of computation in the network; this process helps to limit overfitting of the data.

Figure 2. By a skip connection (i.e. identity mapping) and element-wise addition, a residual block makes it easier to train a deeper network without extra parameter and computational complexity.

Figure 3. Generative adversarial networks (GAN)s consist of a generator and a discriminator, wherein the generator aims to generate sample (synthetic images) that resemble those in the training data while the discriminator tries to distinguish between the two.

Figure 4. A model for detecting five kinds of pulmonary abnormalities (including nodule (ND), consolidation (CS), interstitial opacity (IO), pleural effusion (PE), and pneumothorax (PT)) on chest radiograh with weak labeled data, which indicate the presence of abnormalities' labels only. Upper rows of each case depict the regions of interest labeled by radiologists, and lower rows show the class activation map (CAM). Localization of trained abnormal patterns through CAM could assist radiologists to diagnose lung diseases much easier.

Figure 5. Detection of multiple abnormal lesions on chest radiograh. A CNN model was trained with strongly labeled data, which indicates not only the type of abnormalities but also their locations and boundaries. Multiple lesions were detected in the whole lung

images, and detected regions matched with those of interest delineated by radiologists. This approach could assist radiologists to diagnose and monitor multiple lesions of whole lung.

Figure 6. A method for detecting multi-scale nodules. By Training with RGB color images that were comprised of three adjacent slice in the axial plane, we detected nodules of various sizes. (a)-(d) depict the detected nodules with sizes of 3.4 mm, 5.6 mm, 9.8 mm, and 14.4 mm, respectively.

Figure 7. Regional image patterns of Diffuse interstitial Lung Disease (DILD) using 3D CNN. Since the diagnosis of DILD shows significant variation in inter- and intra-observer interpretation due to a lack of standard criteria and a burden of reviewing a large amount of data, CNN based automated classification on voxel-by-voxel basis is necessary for the quantification of disease extent and distribution of DILD.

Figure 8. We employed 3D U-Net (a typical type of 3D CNN) to develop a robust lobe segmentation. This approach also performed well in the fake and incomplete fissures, since this network was trained on lobe-by-lobe expert human training set.

Figure 9. The fully automated airway segmentation method in a patient with chronic obstructive lung disease (a), which started from (b) the initial airways by using the region growing method. (c) Our method achieved a high sensitivity at a low false positive rate with fast execution time (2-8 min). (d) Manual segmentation usually required 1-2 hours by an experienced research assistant.

Figure 10. Conversion of CT images reconstructed with one kernel to images with different kernels without using a sinogram: (a) CNN architecture for CT kernel conversion. (b)-(c) CT images reconstructed with B10f and B70f, respectively. (d) A CT image interconverted from B10f to B70f using (a). (e)-(f) Difference images between (b)-(c) and (b)-(d), respectively.

Fig 1a

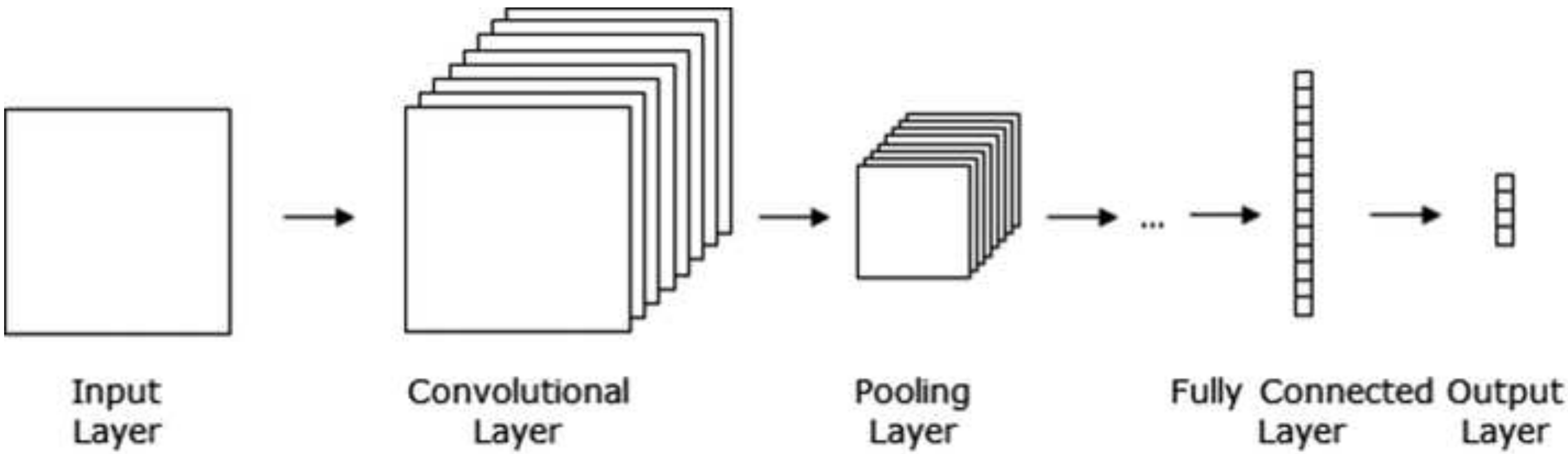

Input
Layer
Convolutional
Layer
Pooling
Layer
...
Fully Connected
Layer
Output
Layer

Fig 1b

Click here to access/download;Figure;Figure 1b.tif

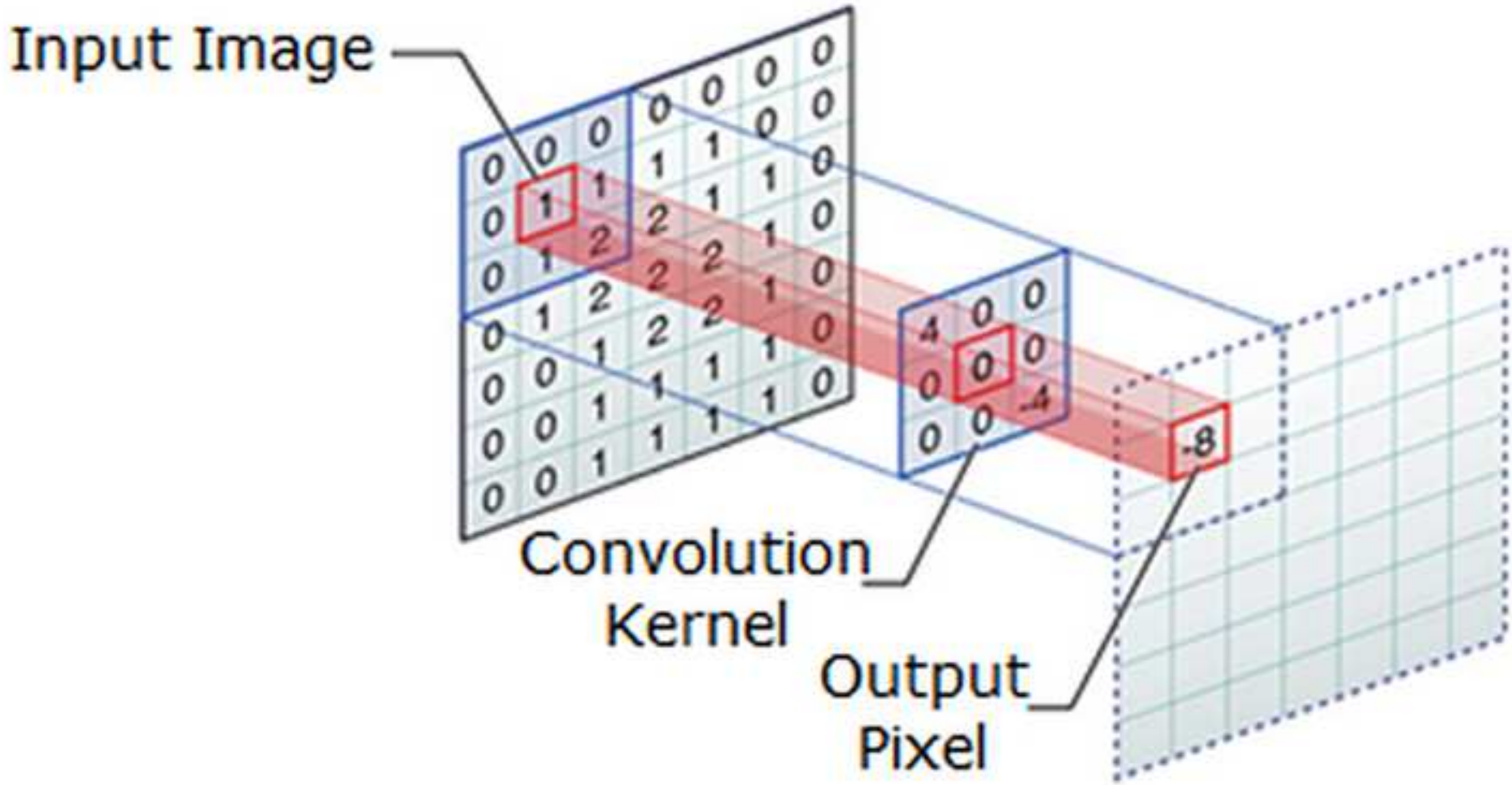
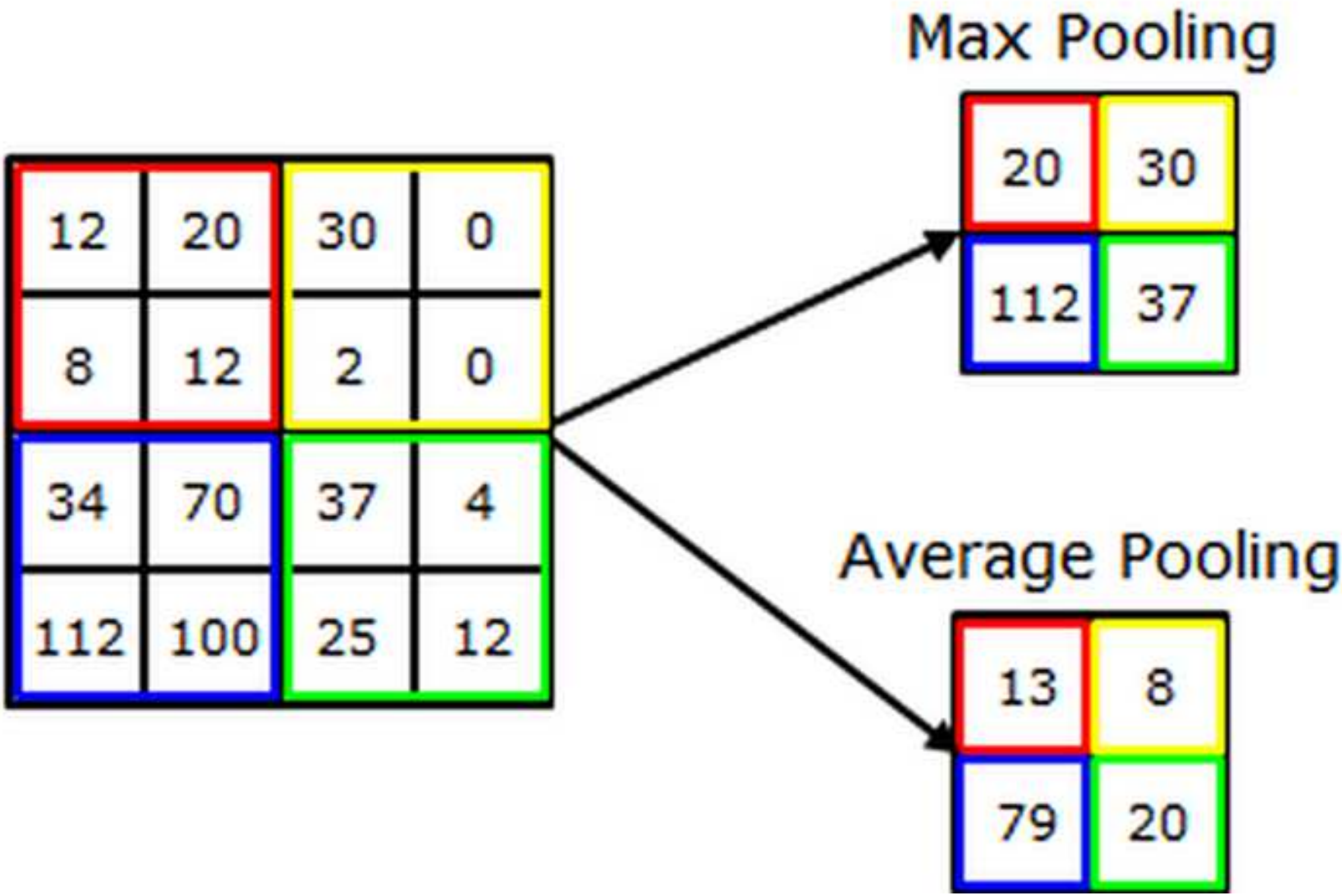


Fig 1c



12
20
30
0
8
12
2
0
34
70
37
4
112
100
25
12
Max Pooling
20
30
112
37
Average Pooling
13
8
79
20

Fig 2

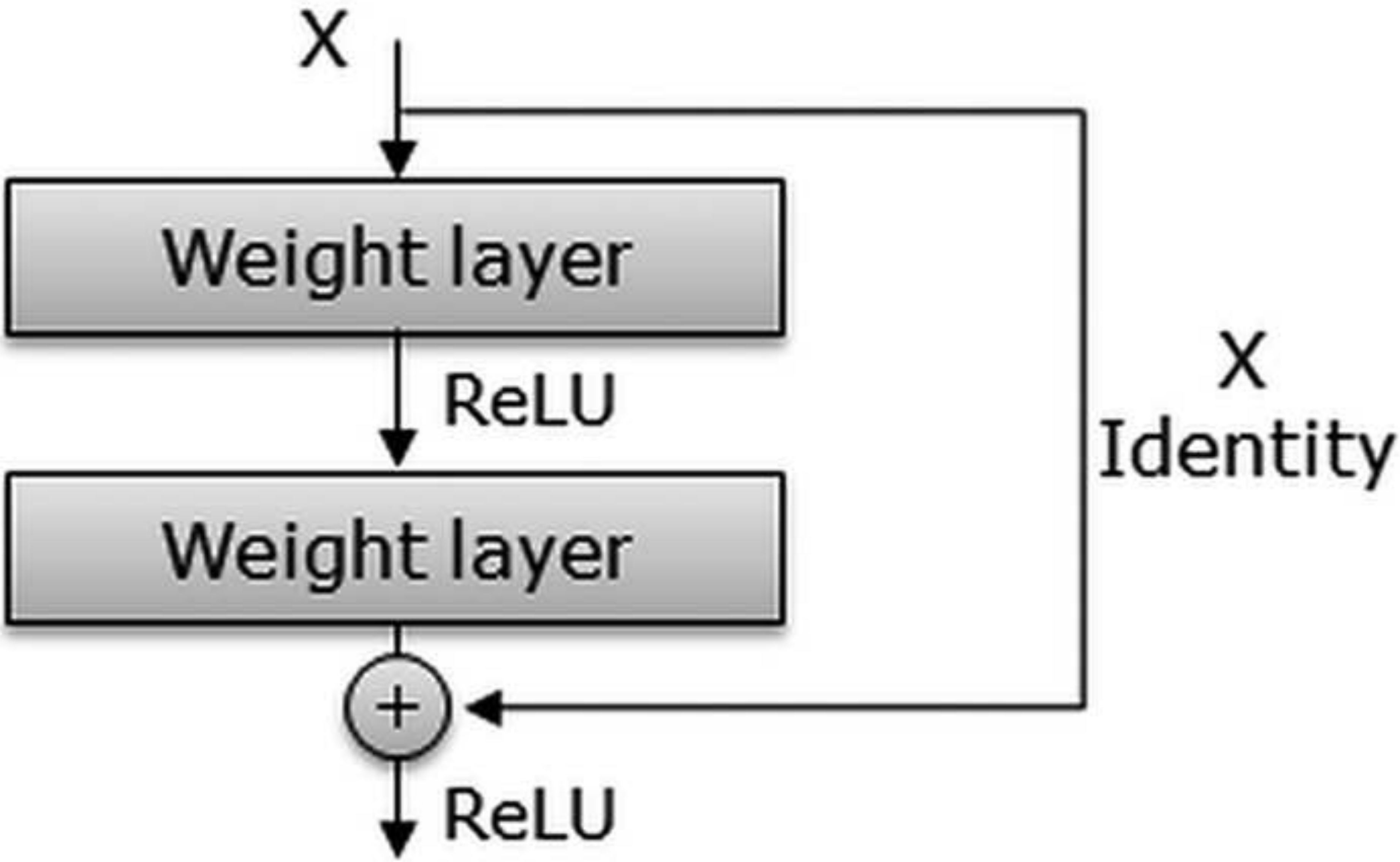
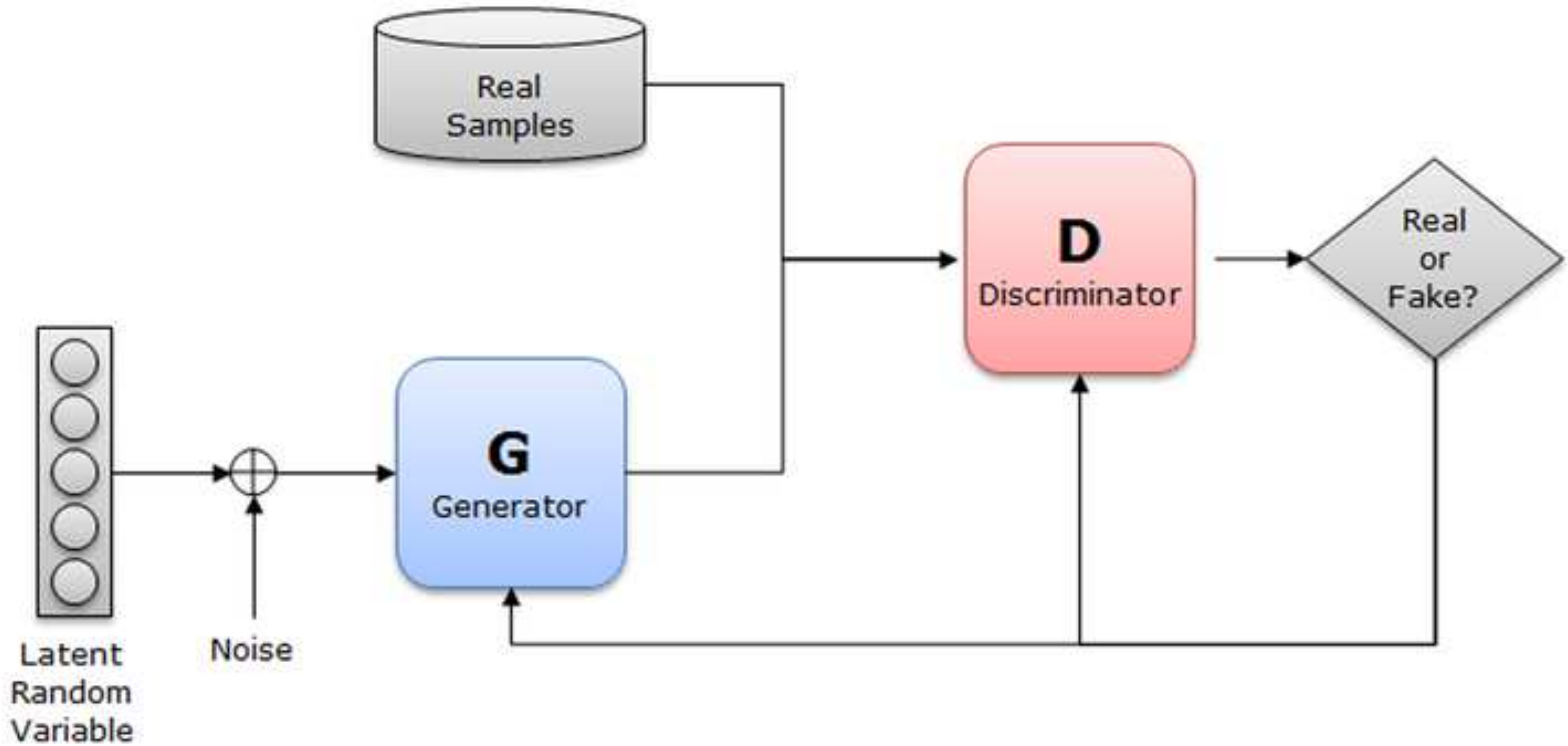


Fig 3


Real
Samples
D
Discriminator
Real
or
Fake?
G
Generator
Latent
Random
Variable
Noise

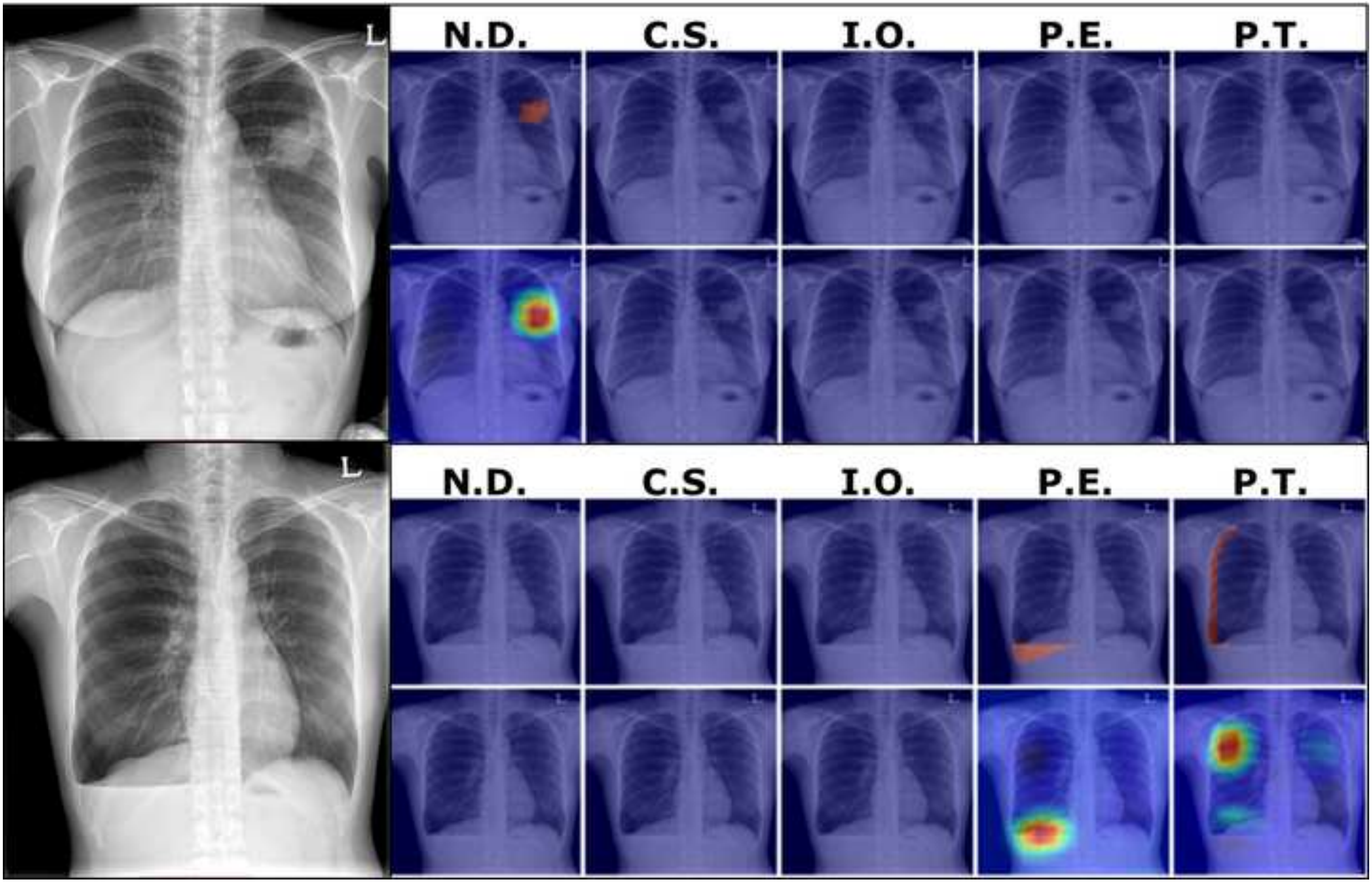

N.D.
C.S.
I.O.
P.E.
P.T.
L
N.D.
C.S.
I.O.
P.E.
P.T.
L

Fig 5

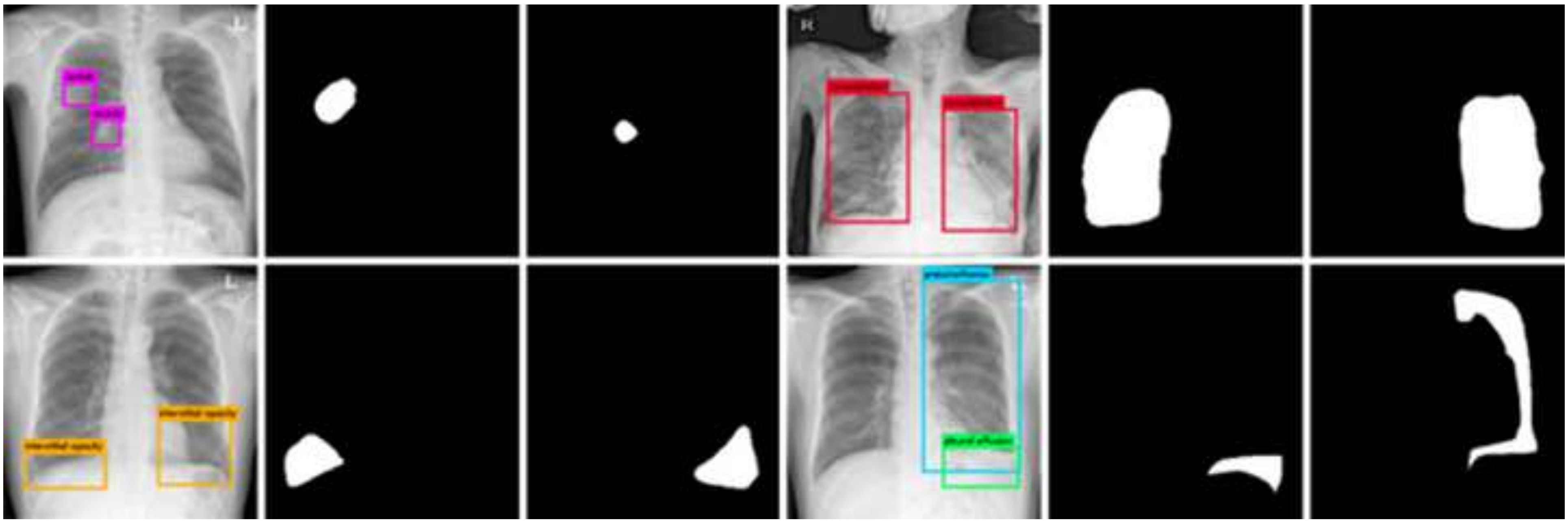


Fig 6a

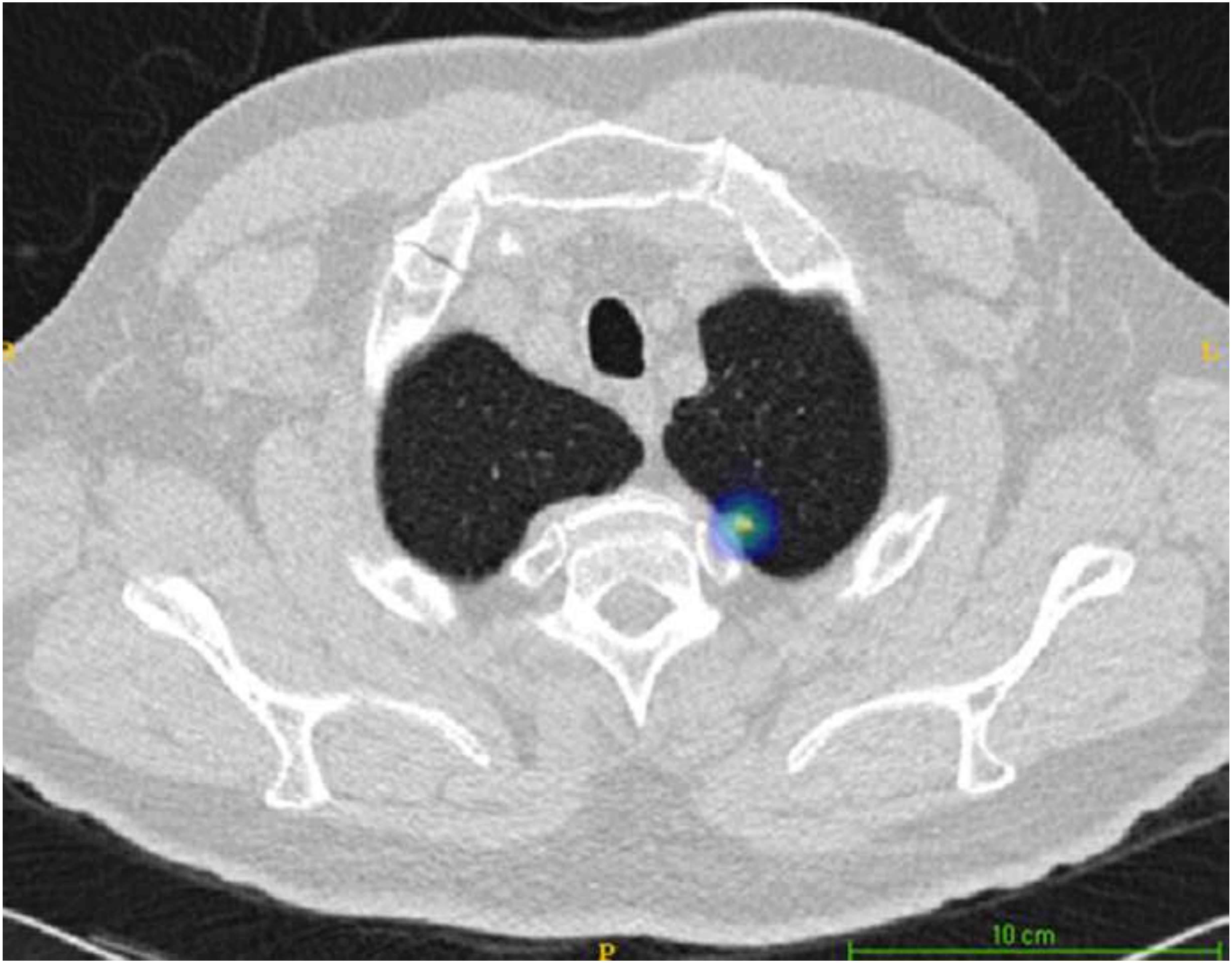


Fig 6b

Click here to access/download;Figure;Figure 6b.tif

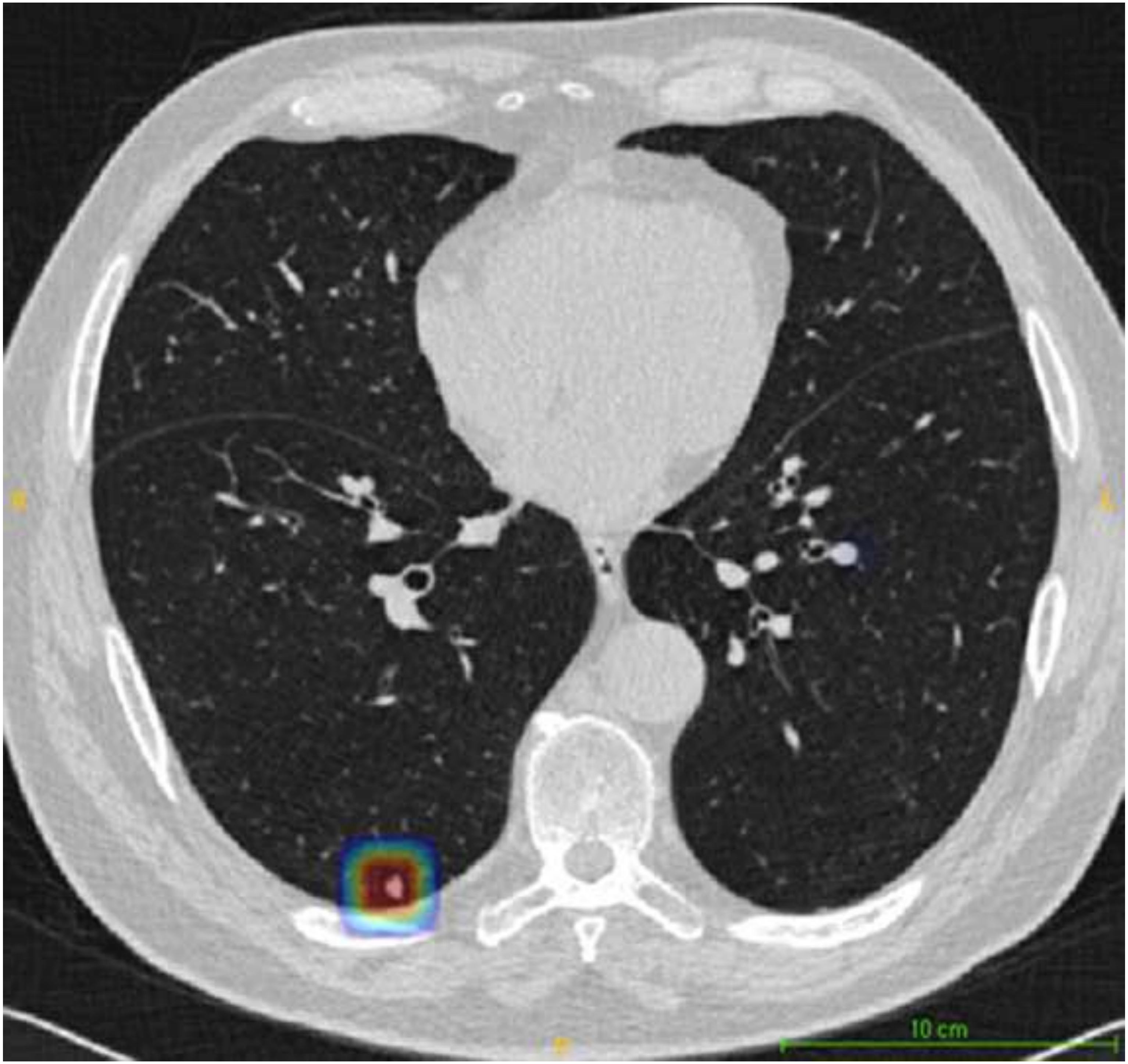


Fig 6c

Click here to access/download;Figure;Figure 6c.tif

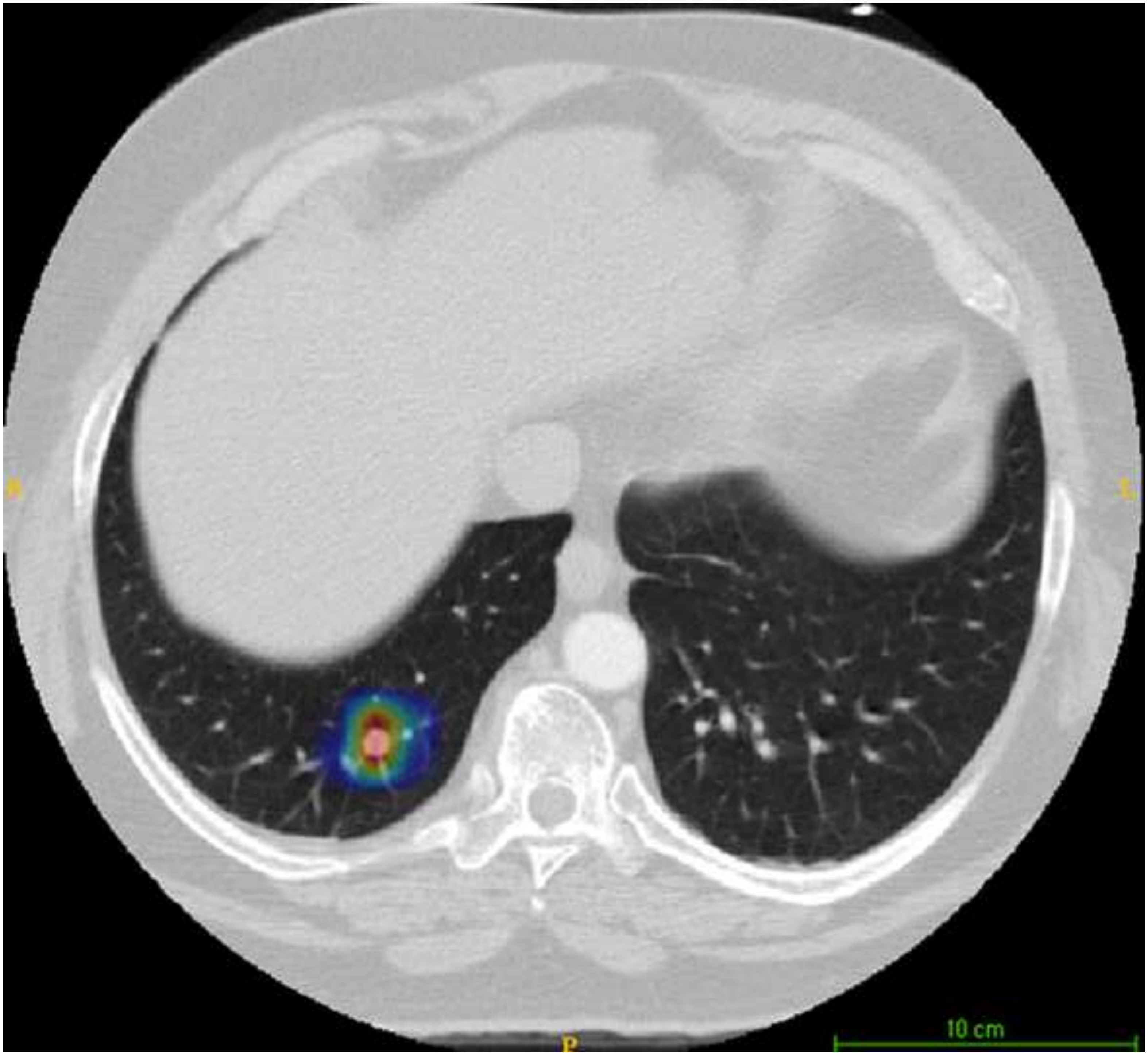


Fig 6d

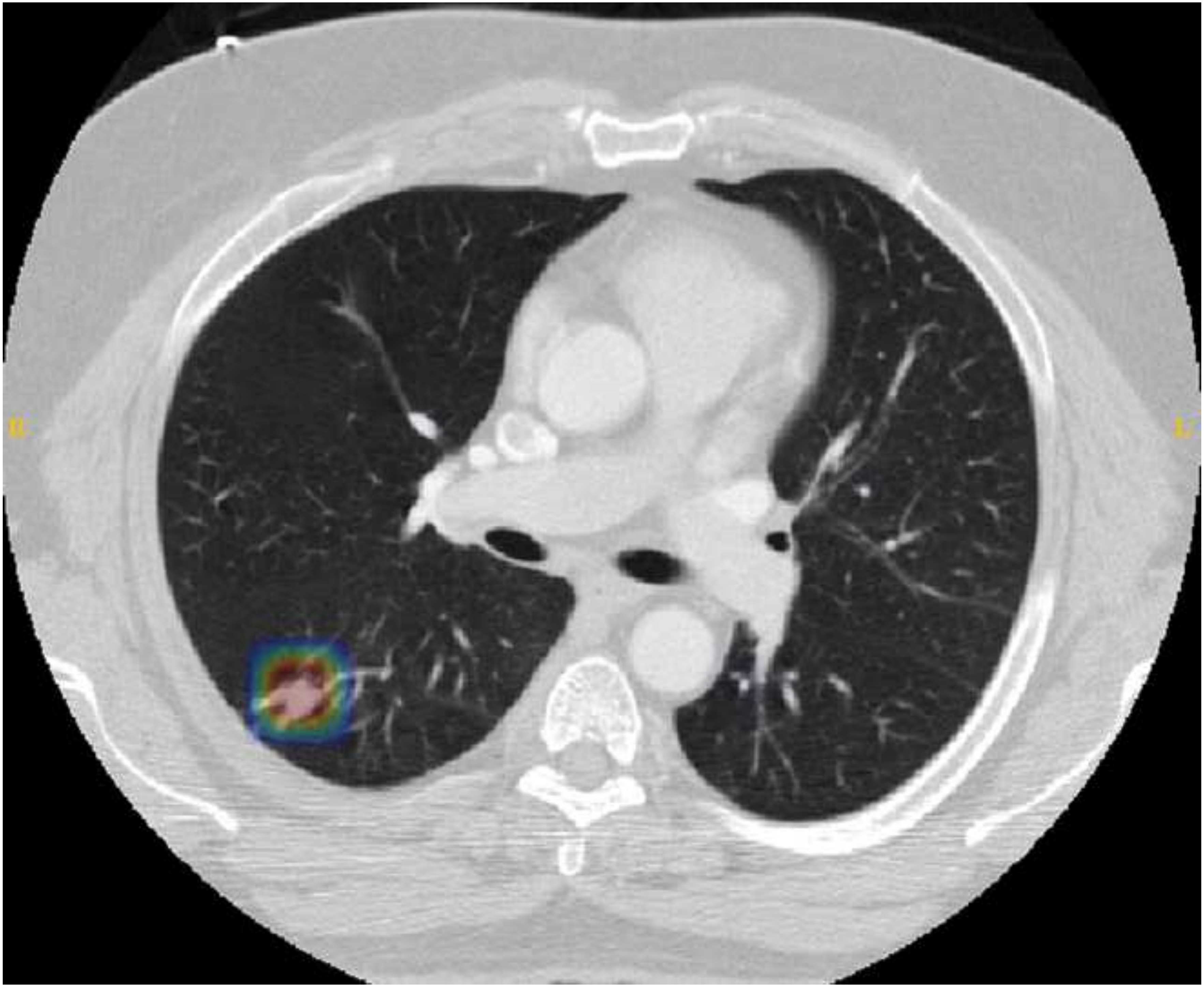


Fig 7

Click here to access/download;Figure;Figure 7.tif

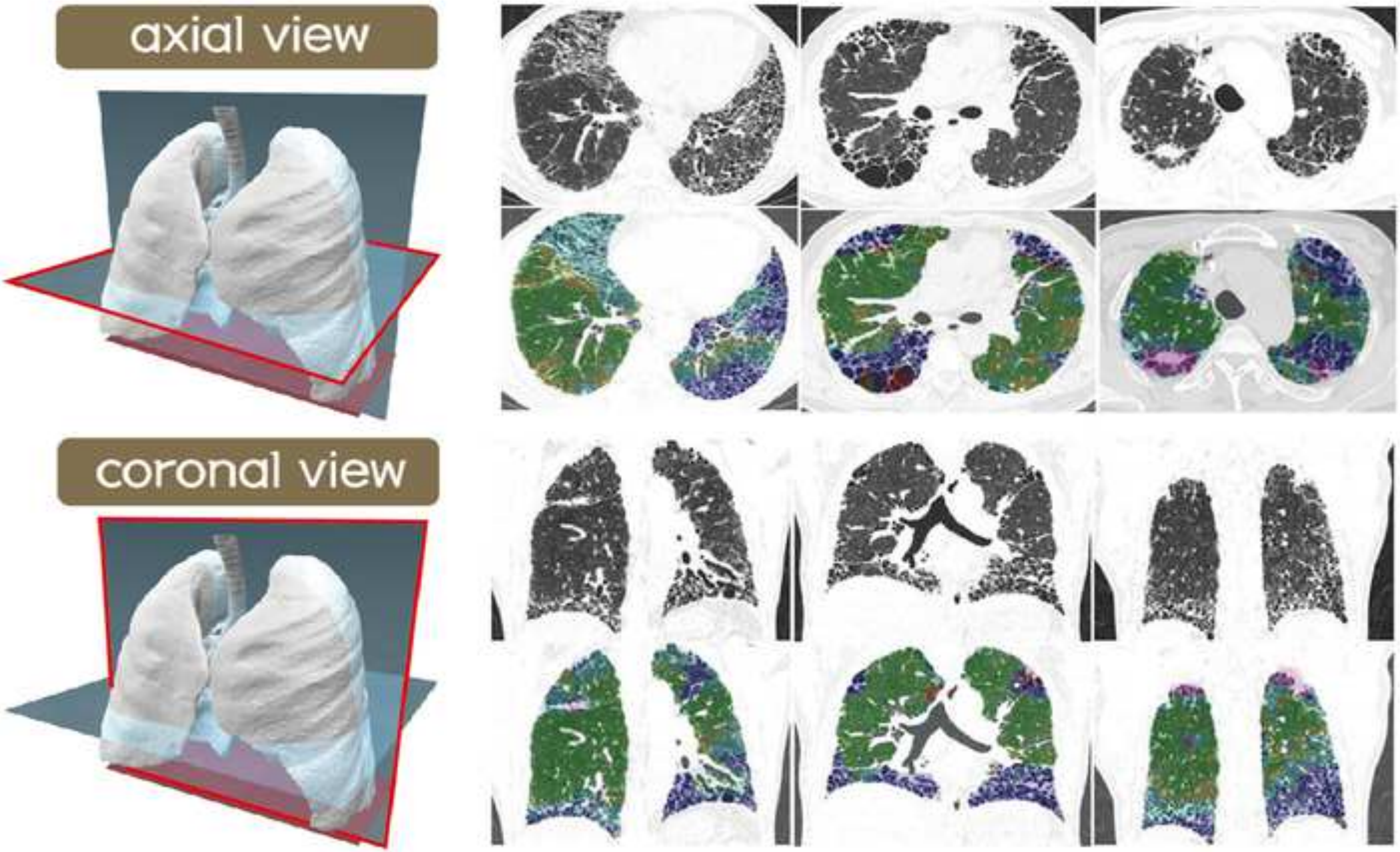


Fig 8

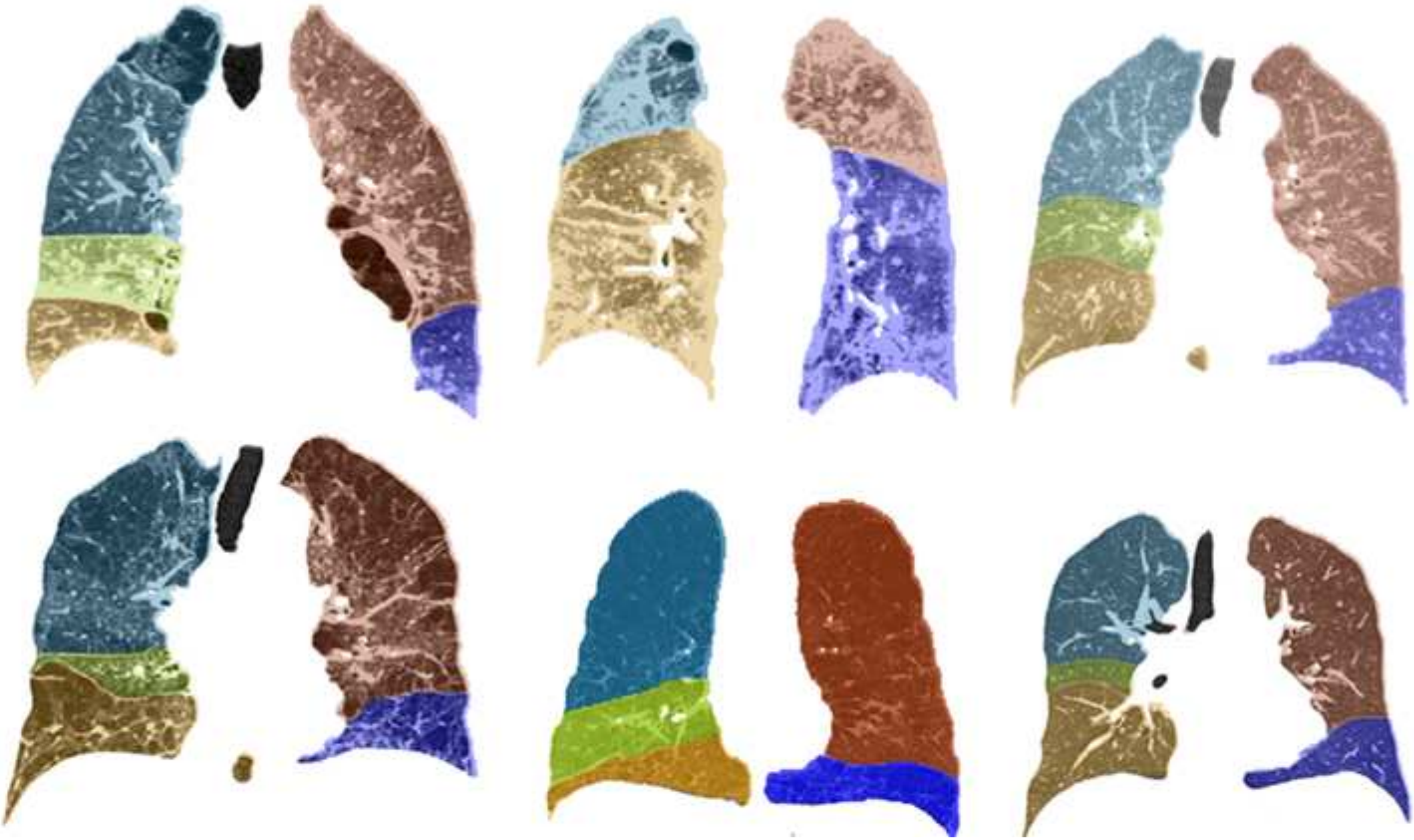
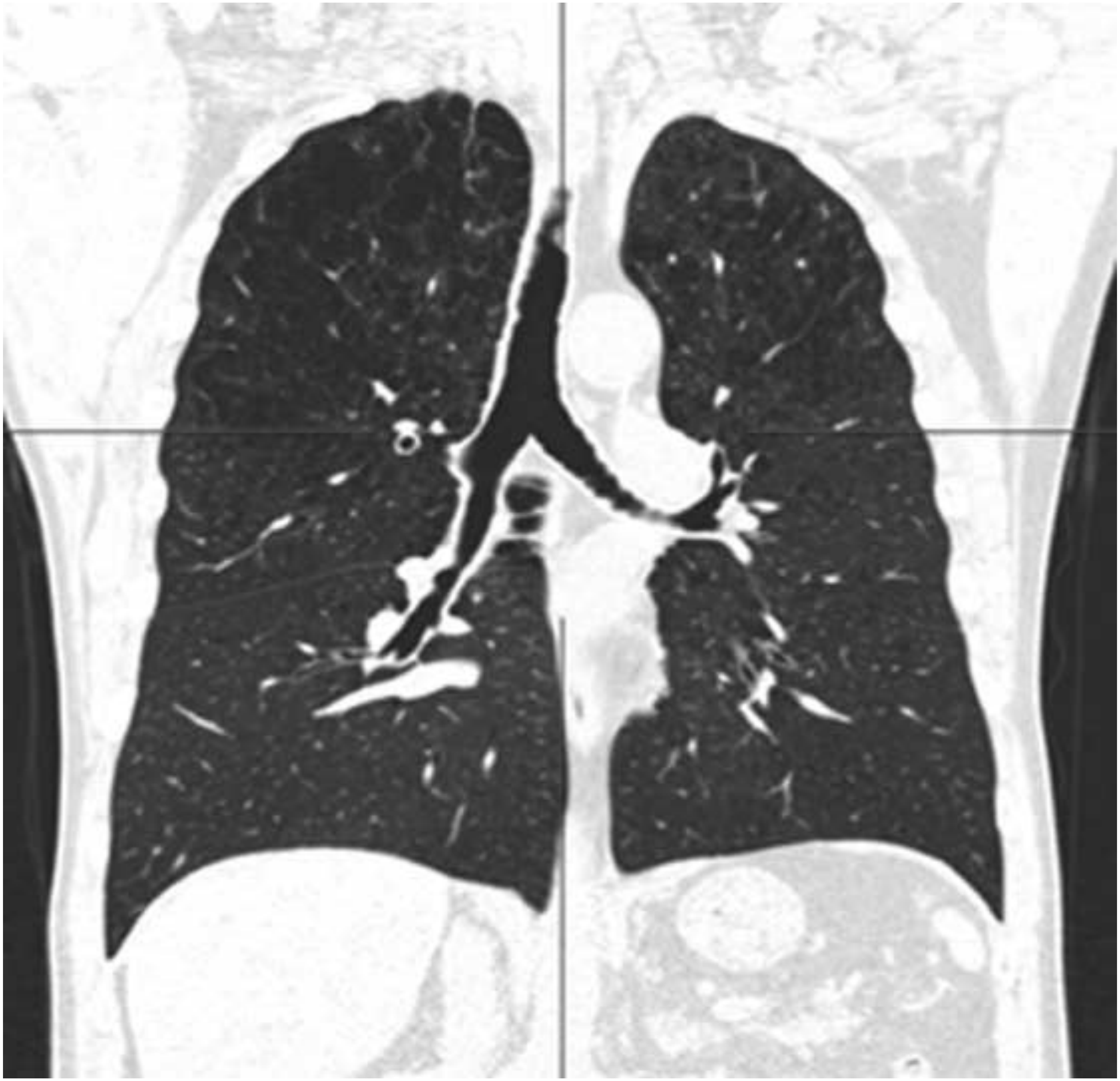


Fig 9a



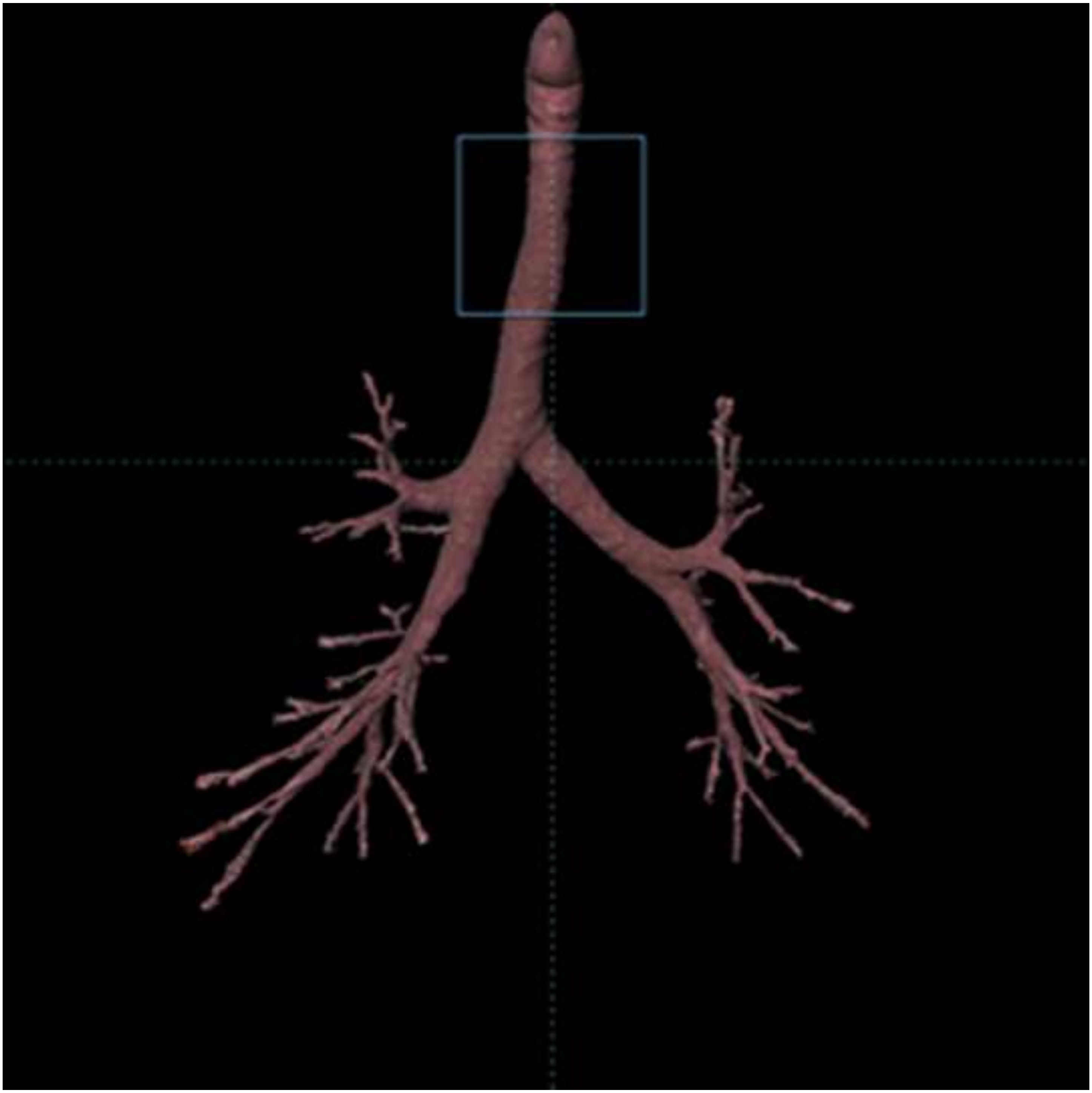
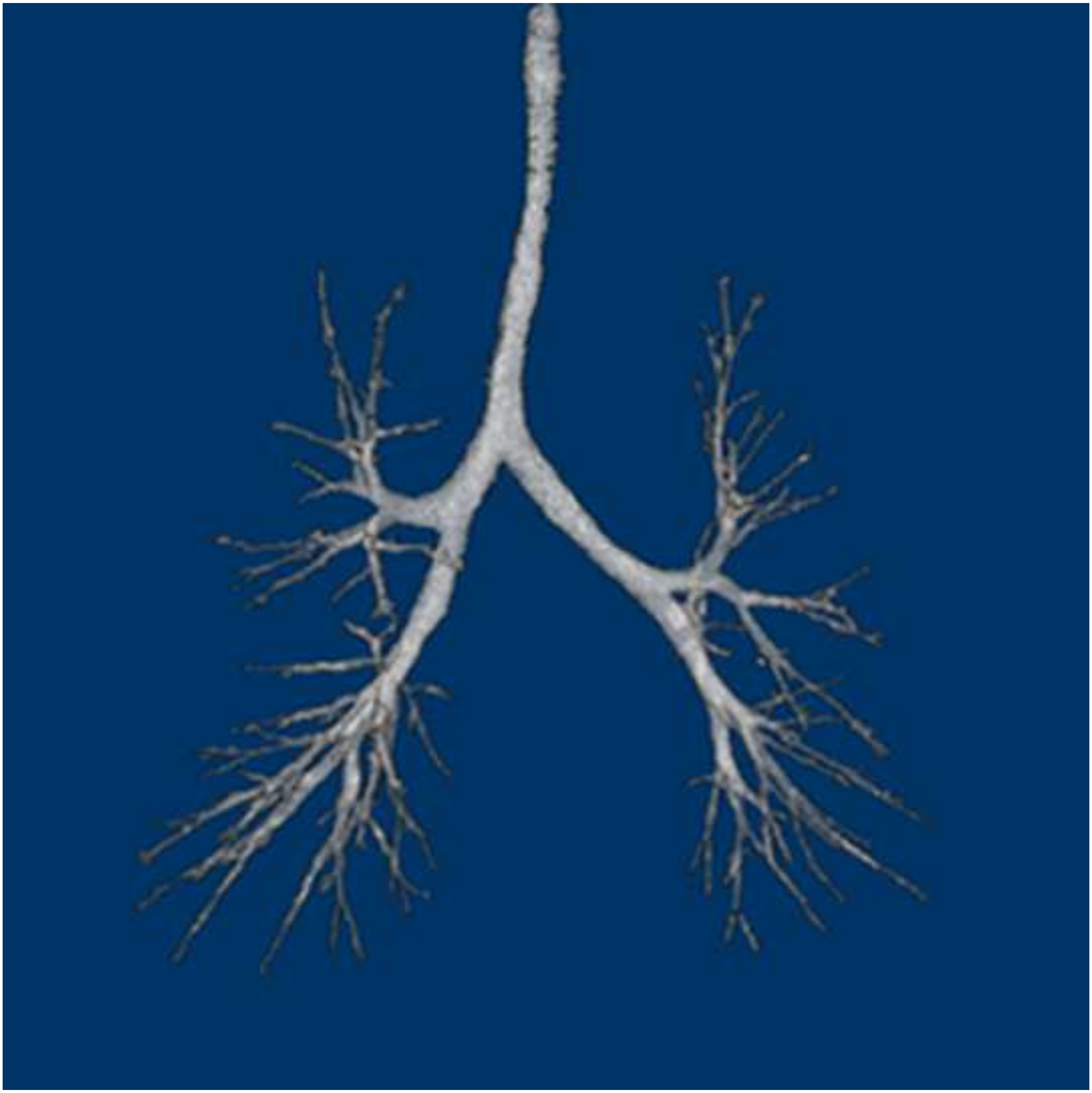


Fig 9c



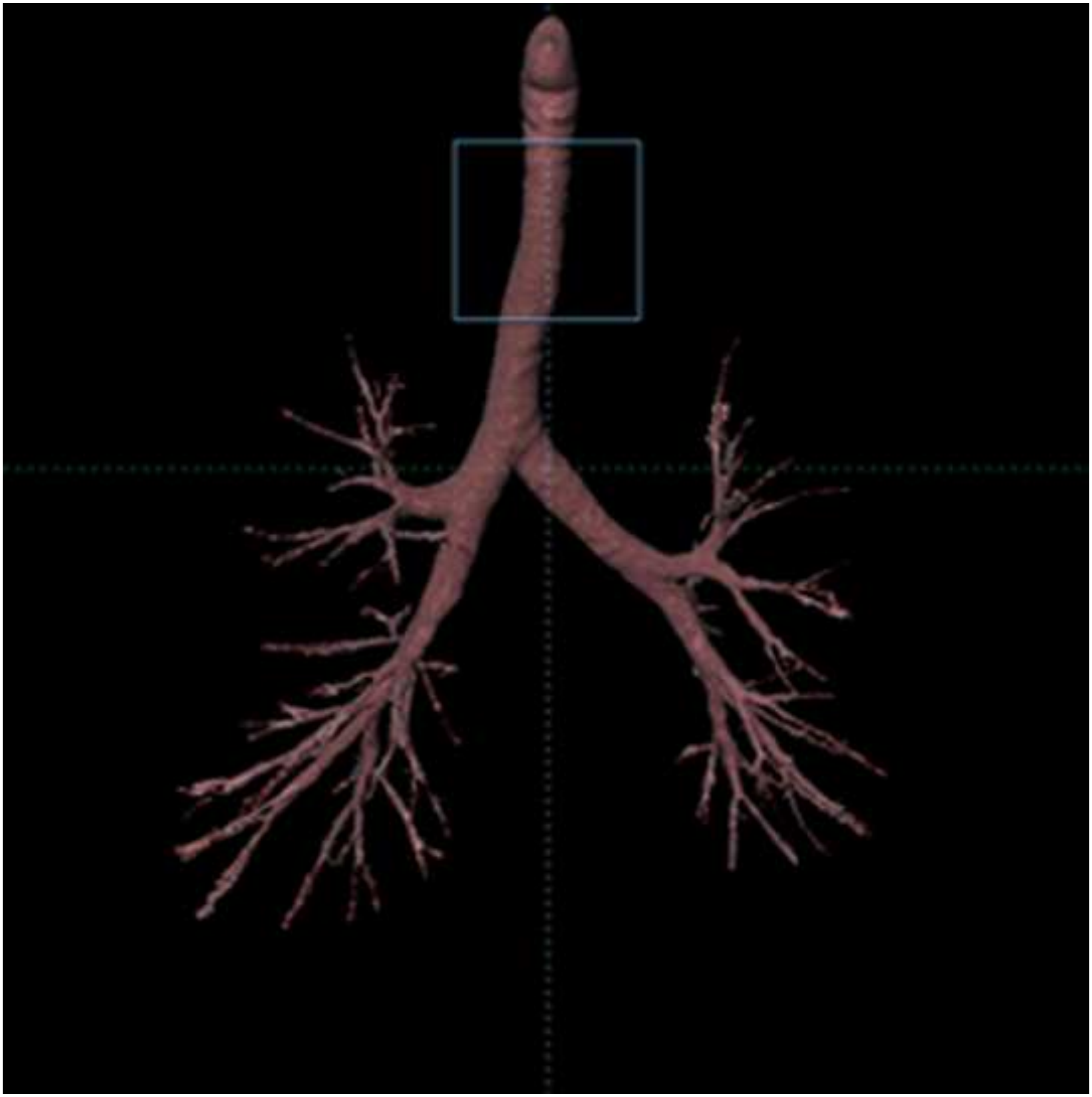


Fig 10a

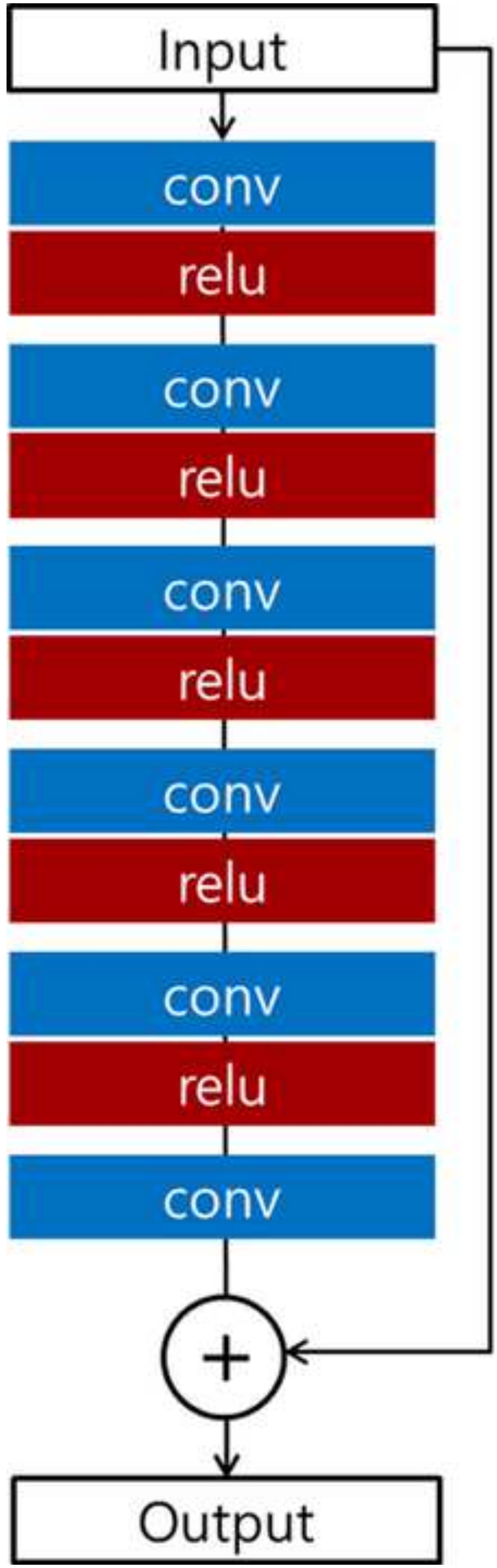

Input
conv
relu
conv
relu
conv
relu
conv
relu
conv
relu
conv
+
Output

Fig 10b

Click here to access/download;Figure;Figure 10b.tif

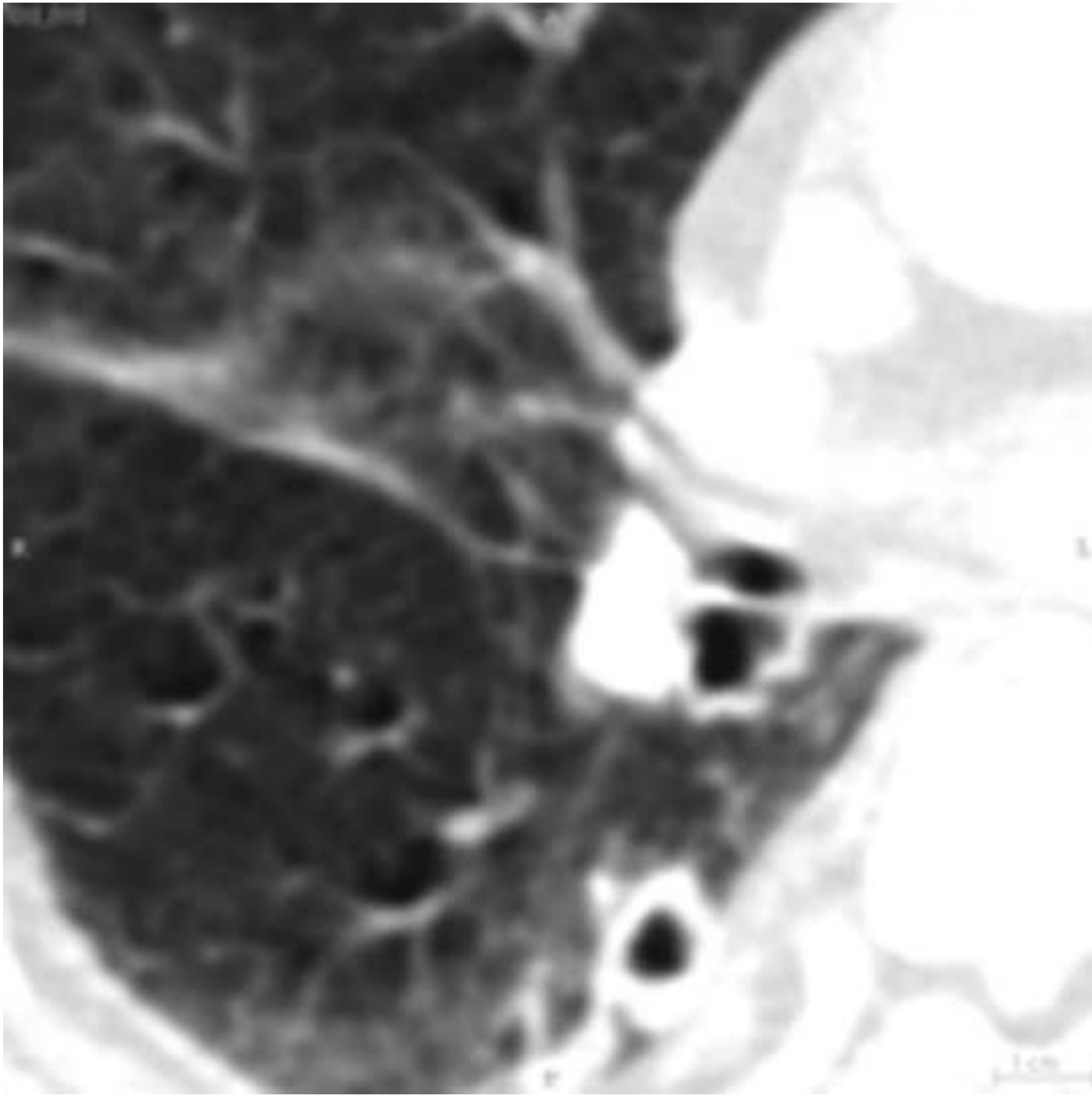


Fig 10c

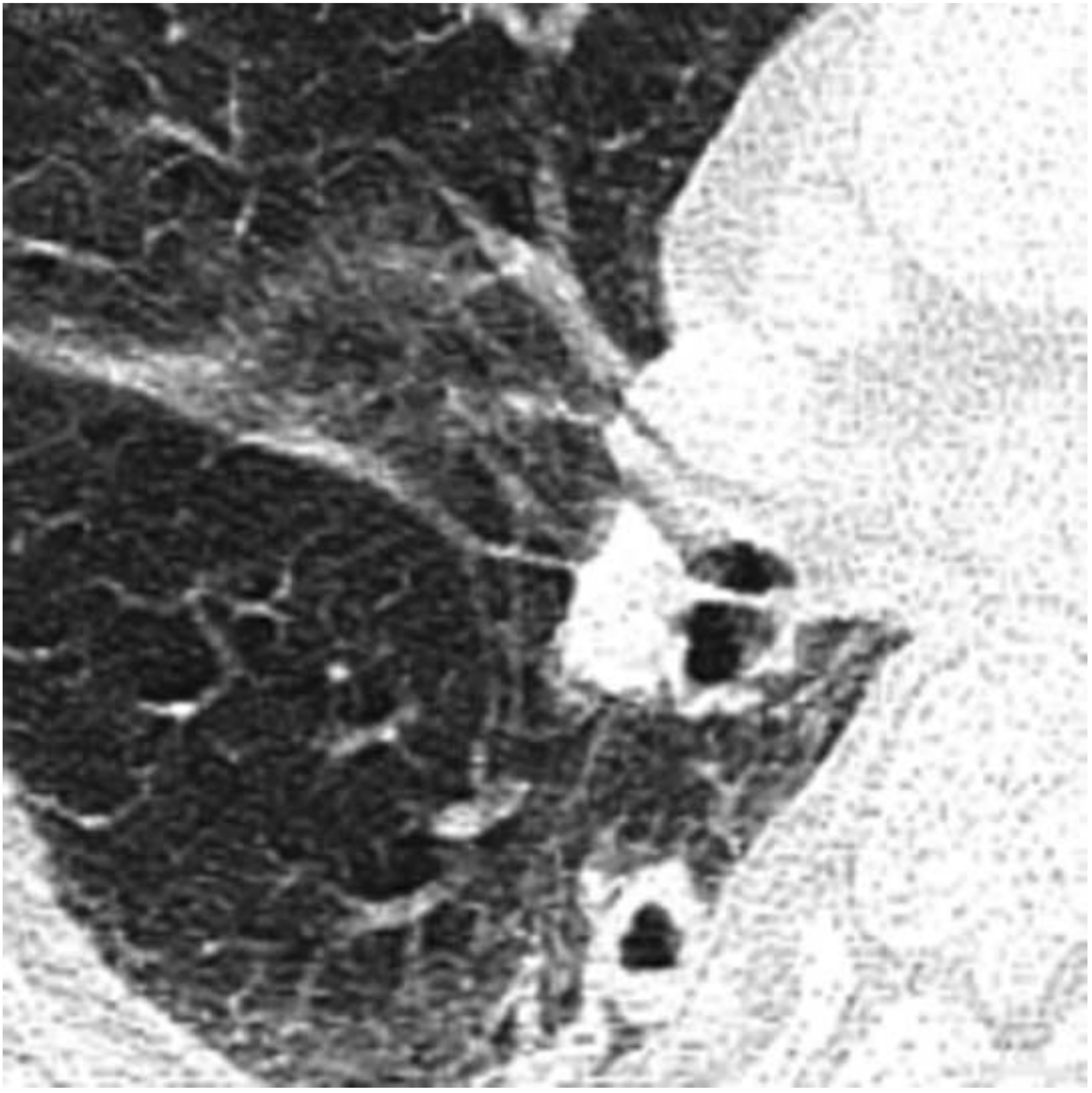


Fig 10d

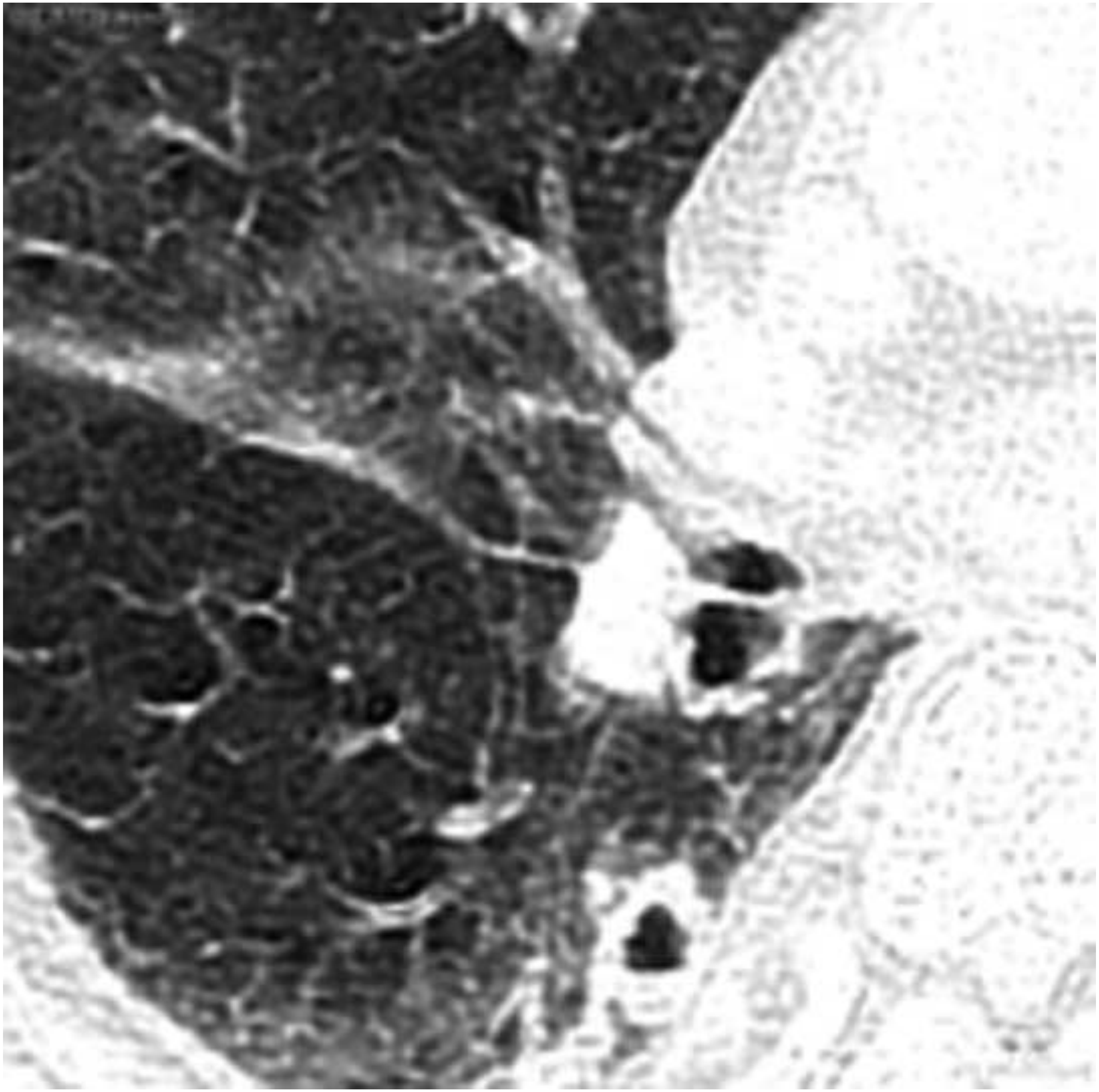


Fig 10e

Click here to access/download;Figure;Figure 10e.tif

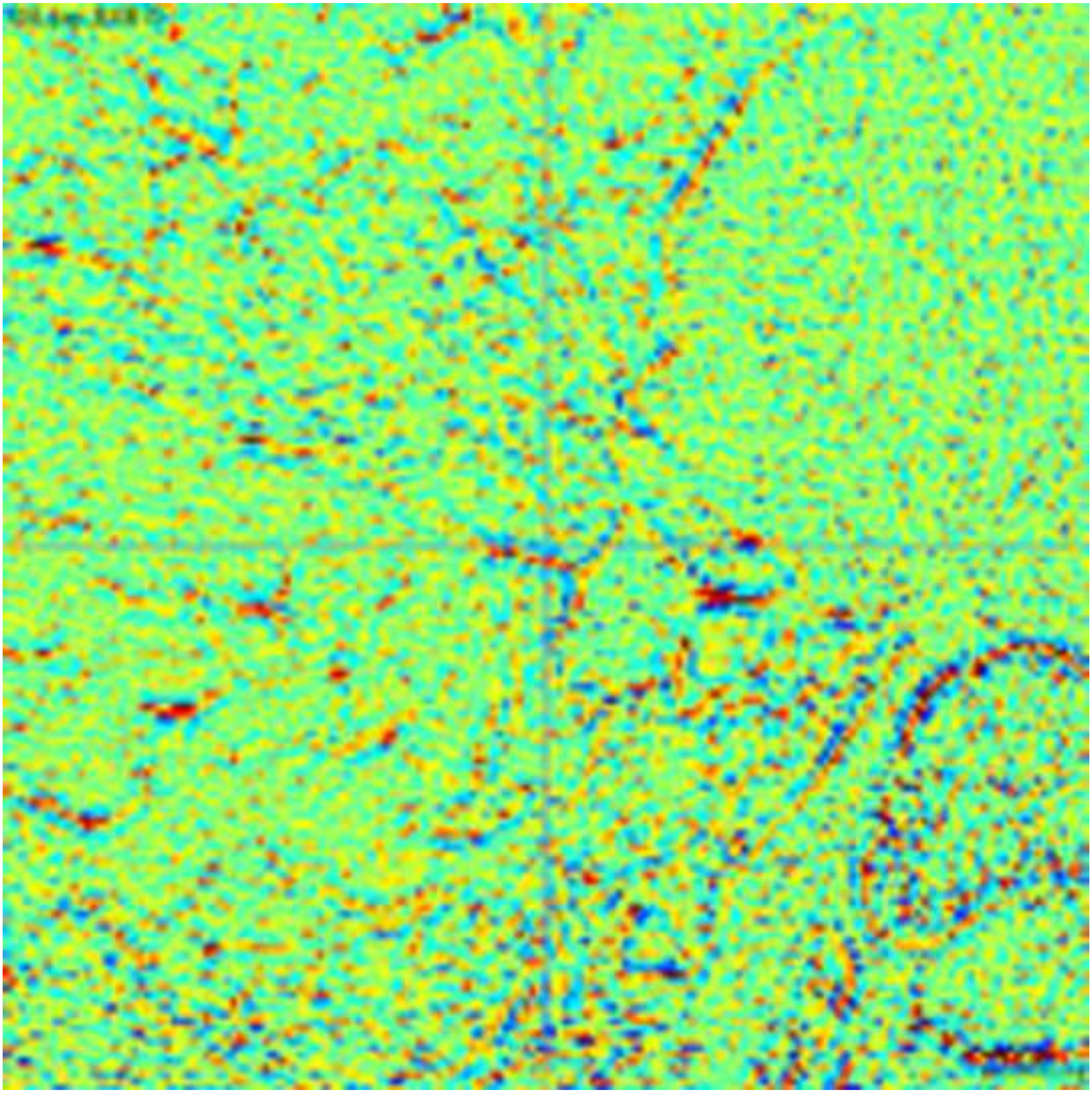


Fig 10f

Click here to access/download;Figure;Figure 10f.tif